

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:50:27 ; Search time 70 Seconds

(without alignments)
612.953 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670

Sequence: 1 MTVTIAINSQNKQIKRLGL.....LFEFRISRNQYVDPFLVTK 322

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

1: A_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1670	100.0	322	22	AAB20106
2	1666	99.8	322	22	AAB20107
3	209	12.5	345	21	AAB21225
4	203	12.2	414	21	AAV74313
5	200	12.0	337	21	AAV74311
6	198	11.9	351	21	AAV74312
7	158.5	9.5	436	23	ABR48721
8	151.5	9.1	478	14	AAK45178
9	146.5	8.8	439	23	ABB53578
10	146.5	8.8	482	23	ABB48659

11	146.5	8.8	484	16	AAK73913
12	134	8.0	410	23	AAO17658
13	131	7.8	410	23	AAO17656
14	130.5	7.8	437	17	AAK85285
15	130.5	7.8	518	22	ABB06928
16	130.5	7.8	894	23	ABB54367
17	130	7.8	410	20	AAV13394
18	127	7.6	1708	20	AAV13394
19	126.5	7.6	374	13	AAK24147
20	126.5	7.6	374	21	AAV44646
21	125.5	7.5	205	22	AAK92487
22	124	7.4	288	20	AAW95501
23	123.5	7.4	666	17	AAK85291
24	120.5	7.2	639	22	AAK94053
25	120	7.2	357	23	AAK18364
26	120	7.2	382	23	AAK18360
27	119.5	7.2	1426	20	AAK13432
28	119.5	7.2	1426	23	AAK16323
29	118	7.1	216	17	AAK85287
30	118	7.1	284	21	AAV95030
31	118	7.1	633	22	AAK52321
32	118	7.1	633	22	AAK57365
33	118	7.1	1192	21	AAK56967
34	118	7.1	1192	22	AAU04591
35	118	7.1	1192	22	AAK82349
36	118	7.1	1246	22	AAU33228
37	118	7.1	1708	20	AAW93408
38	117.5	7.0	364	23	AAK18363
39	117.5	7.0	389	23	ABP25889
40	117.5	7.0	389	23	AAK18359
41	117	7.0	237	22	AAK90704
42	115.5	6.9	576	22	ABB61241
43	115	6.9	1624	22	ABB60238
44	113.5	6.8	202	22	ABB61097
45	112	6.7	275	21	AAK55813

ALIGNMENTS

RESULT 1
AAB20106
ID AAB20106 standard; Protein; 322 AA.
XX AC AAB20106;
XX DT 23-APR-2001 (first entry)
XX DE Moraxella catarrhalis BASB110 protein.
XX KW BASB110; infection; otitis media; pneumonia; therapy; diagnosis;
XX KW antibacterial; antimicrobial; vaccine.
XX OS Moraxella catarrhalis.
XX FT Key location/Qualifiers
XX FT Misc-difference 104 /note= "Gly in translation of BASB110 PCR product"
XX PD 04-JAN-2001.
XX PF 23-JUN-2000; 2000WO-EP05854.
XX PR 25-JUN-1999; 99GB-0015031.
XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Thomard J;
XX WP1; 2001-112459/12.
XX DR N-PSDB; AAF30046.

Listeria monocytog
H influenzae BASB2
H influenzae BASB2
Lysin. Lactococcus
Micrococcus car
Lactococcus lactis
H influenzae BASB2
HEV-US2 ORF1 prote
Beta-lytic proteas
Achromacter lyti
C glutamicum p
B. subtilis Yoch r
Enterococcus hirae
Human protein sequ
Streptococcus pyog
Streptococcus pyog
Truncated cellulase
Active cellulase p
Lysin carboxy term
Human clone vb22.1
Las17 protein. Sa
Amino acid sequenc
Human MAGI polypep
Human Nogo protein
Human Nogo-A prote
Novel human secret
Swine HEN ORF 1 pr
Streptococcus pyog
Streptococcus pyog
Streptococcus pyog
C glutamicum prote
Drosophila melanog
Drosophila melanog
Rat nerve growth f

XX Novel BASH10 polypeptides of Moraxella catarrhalis, useful as a
 PT vaccine for treating Moraxella catarrhalis infections -
 PS Claim 1; Page 82-83; 88pp; English.

CC The present sequence is that of BASH10 protein from Moraxella
 CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
 CC media in children and pneumonia in adults. The sequence is a
 CC translation of the BASH10 gene coding region (see AAF30047). It
 CC differs by 1 residue from the predicted polypeptide sequence (see
 CC AAB20107) of a BASH10 PCR product, having Ser rather than Gly at
 CC position 104. The invention provides BASH10 polypeptides, and at
 CC polynucleotides encoding them, as well as expression vectors, host
 CC cells and methods for producing BASH10 polypeptides using
 CC recombinant methods. Also claimed is a vaccine composition
 CC comprising a BASH10 polypeptide, an immunogenic fragment of a
 CC acid sequence identity to BASH10, or comprising a polynucleotide
 CC encoding such a polypeptide. A claimed method of diagnosing a
 CC Moraxella infection involves identifying a BASH10 polypeptide or
 CC antibody. A claimed therapeutic composition useful in treating
 CC humans with M. catarrhalis infection comprises at least 1 antibody
 CC directed against a BASH10 polypeptide. BASH10 polypeptides also
 CC have utility in raising specific antibodies, and in screening for
 CC antibacterial drugs.

XX Sequence 322 AA;
 SQ

Query Match 100.0%; Score 1670; DB 22; Length 322;
 Best Local Similarity 100.0%; Pred. No. 5.2e-134;
 Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTVTIANSONOKPIRGLIRGVITTTCTILAGCAKPTVNTSGSGSHRTSGGGLAIGS 60
 DB 1 MTVTIANSONOKPIRGLIRGVITTTCTILAGCAKPTVNTSGSGSHRTSGGGLAIGS 60
 QY 61 QVITDSQGVNRYOVKQDPTVSKIAORGLNWRIGHINNLSYITYGWLTMSGDL 120
 DB 61 QVITDSQGVNRYOVKQDPTVSKIAORGLNWRIGHINNLSYITYGWLTMSGDL 120
 QY 121 KVERSISSGVNTATTPSPVAVOSSRPVVOQHPAVOKPTPPVVKPPPPVVOQPP 180
 DB 121 KVERSISSGVNTATTPSPVAVOSSRPVVOQHPAVOKPTPPVVKPPPPVVOQPP 180
 QY 181 VAPPTTEAPFATGSSGMOPRIPVAGTNPVRRFGTATVAGSTVNSGMMFSGRGGDLIN 240
 DB 181 VAPPTTEAPFATGSSGMOPRIPVAGTNPVRRFGTATVAGSTVNSGMMFSGRGGDLIN 240
 QY 241 ASNAGTVIOADHNMDDASIVIOHTNGFVSSYTHIKDAQVKTGDPVTRGRIASMKQPSG 300
 DB 241 ASNAGTVIOADHNMDDASIVIOHTNGFVSSYTHIKDAQVKTGDPVTRGRIASMKQPSG 300
 QY 301 AALFEFRISNGVYVDPPLTVLK 322
 DB 301 AALFEFRISNGVYVDPPLTVLK 322

RESULT 2

AA20107
 ID AAB20107 standard; Protein, 322 AA.
 AC AAB20107;

XX 23-APR-2001 (first entry)
 XX Moraxella catarrhalis BASH10 protein.
 XX BASH10; infection; otitis media; pneumonia; therapy; diagnosis;
 XX antibacterial; antimicrobial; vaccine.
 XX Moraxella catarrhalis.

FF Key Location/Qualifiers
 FT MISC-difference 104 /note= "Ser in translation of BASH10 gene"
 PS WO200100838-A1.
 XX 04-JAN-2001.
 XX 23-JUN-2000; 2000WO-BP05854.
 XX 25-JUN-1999; 99GB-0015031.
 XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Thonard J;
 PI WPI, 2001-112459/12.
 DR N-PDB; AAF30047.

XX Novel BASH10 polypeptides of Moraxella catarrhalis, useful as a
 PT vaccine for treating Moraxella catarrhalis infections -
 PS Claim 1; Page 83-84; 88pp; English.

CC The present sequence is that of BASH10 protein from Moraxella
 CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
 CC media in children and pneumonia in adults. The sequence is a
 CC translation of a BASH10 PCR product (see AAF30047). It differs by
 CC 1 residue from the predicted polypeptide sequence (see AAB20106) of
 CC the BASH10 gene product, having Gly rather than Ser at position
 CC 104. The invention provides BASH10 polypeptides, and
 CC polynucleotides encoding them, as well as expression vectors, host
 CC cells and methods for producing BASH10 polypeptides using
 CC recombinant methods. Also claimed is a vaccine composition
 CC comprising a BASH10 polypeptide, an immunogenic fragment of a
 CC acid sequence identity to BASH10, or comprising a polynucleotide
 CC encoding such a polypeptide. A claimed method of diagnosing a
 CC Moraxella infection involves identifying a BASH10 polypeptide or
 CC antibody. A claimed therapeutic composition useful in treating
 CC humans with M. catarrhalis infection comprises at least 1 antibody
 CC directed against a BASH10 polypeptide. BASH10 polypeptides also
 CC have utility in raising specific antibodies, and in screening for
 CC antibacterial drugs.

XX Sequence 322 AA;
 SQ

Query Match 99.8%; Score 1666; DB 22; Length 322;
 Best Local Similarity 99.7%; Pred. No. 1.1e-133;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTVTIANSONOKPIRGLIRGVITTTCTILAGCAKPTVNTSGSGSHRTSGGGLAIGS 60
 DB 1 MTVTIANSONOKPIRGLIRGVITTTCTILAGCAKPTVNTSGSGSHRTSGGGLAIGS 60
 QY 61 QVITDSQGVNRYOVKQDPTVSKIAORGLNWRIGHINNLSYITYGWLTMSGDL 120
 DB 61 QVITDSQGVNRYOVKQDPTVSKIAORGLNWRIGHINNLSYITYGWLTMSGDL 120
 QY 121 KVERSISSGVNTATTPSPVAVOSSRPVVOQHPAVOKPTPPVVKPPPPVVOQPP 180
 DB 121 KVERSISSGVNTATTPSPVAVOSSRPVVOQHPAVOKPTPPVVKPPPPVVOQPP 180
 QY 181 VAPPTTEAPFATGSSGMOPRIPVAGTNPVRRFGTATVAGSTVNSGMMFSGRGGDLIN 240
 DB 181 VAPPTTEAPFATGSSGMOPRIPVAGTNPVRRFGTATVAGSTVNSGMMFSGRGGDLIN 240
 QY 241 ASNAGTVIOADHNMDDASIVIOHTNGFVSSYTHIKDAQVKTGDPVTRGRIASMKQPSG 300
 DB 241 ASNAGTVIOADHNMDDASIVIOHTNGFVSSYTHIKDAQVKTGDPVTRGRIASMKQPSG 300
 QY 301 AALFEFRISNGVYVDPPLTVLK 322
 DB 301 AALFEFRISNGVYVDPPLTVLK 322

```

DB          301  AALFEFRISRNQGVYDPLTLVK 322

RESULT 3
ID  AAB21225
XX  AAB21225 standard; Protein; 345 AA.
AC  AAB21225;
XX
DT  -12-JAN-2001 (first entry)
XX
DB  Haemophilus somnus lipB.
XX
KW  Haemophilus somnus; lipoprotein B; lipB; transferrin-binding protein;
KW  tbp; antibacterial; vaccine; infection;
KW  thromboembolic meningococcalitis; ITMEM; myocarditis; septicemia;
XX  arthritis; pneumonia.
XX  Haemophilus somnus.
XX
EN  WO200053765-A1.
XX
PD  14-SEP-2000.
XX
PE  10-MAR-2000; 2000WO-CA00244.
XX
PR  10-MAR-1999; 99US-0267749.
XX  24-SEP-1999; 99US-0405728.
XX
PA  (UNISA-) UNIV SASKATCHEWAN.
XX
PI  Potter AA, Rioux C, Schryvers AB;
XX
DR  WPI; 2000-565601/52.
XX  N-PSDB; AAA93349.
XX
PT  Nucleic acids encoding Haemophilus somnus transferrin binding proteins
PT  useful for vaccinating against and diagnosing H. somnus infections e.g.
PT  myocarditis and pneumonia in cattle -
XX
PS  Disclosure; Fig 11; 67pp; English.
XX
XX
XX  The present sequence is the Haemophilus somnus lipoprotein lipB.
XX  Recombinant vaccines containing H. somnus lipoprotein lipB, lipB
XX  and lipC have been described. H. somnus transferrin-binding proteins Tbp1
XX  and Tbp2 have also been found to be effective in vaccines against
XX  H. somnus infection. Clones expressing Tbp epitopes were identified by
XX  screening a genomic expression library of H. somnus strain H825 in
XX  E. coli with polyclonal antiserum raised against affinity-purified Tbp1
XX  and Tbp2 of H. somnus. The genes coding for Tbp1 and Tbp2 were obtained
XX  by inverse PCR. The nucleotide sequence encoding Tbp1 and Tbp2 may be
XX  used for the recombinant production of Tbp1 and Tbp2, which may then be
XX  used to manufacture vaccine compositions for immunising against H. somnus
XX  infections. The antibodies raised against the transferrin binding
XX  proteins may also be used diagnostically to identify the presence of
XX  H. somnus infections. H. somnus is a pathological Gram-negative bacteria
XX  that causes a number of diseases in cattle such as thromboembolic
XX  meningococcalitis (ITMEM), myocarditis, septicemia, arthritis and
XX  pneumonia.
XX
SQ  Sequence 345 AA;

Query Match 12.5%; Score 209; DB 21; Length 345;
Best Local Similarity 24.8%; Pred. No. 1.4e-09;
Matches 63; Conservative 49; Mismatches 106; Indels 36; Gaps 6

OY  73 YQVKGQGVSKAQRVGLNWRKIGHINNINSYTYITGQWLTMSG---DLKVERRSIS 129
DB  120 YKVRGDTMFELIAYISGMDIKELATLNNSSEFYHISIGCVLKIANNIPSNMIPROTEIN 179
OY  130 GVNTAHTSPVAVQSSRPVQGNAPVQKTPRVVVVKKKTPRPVVQCAPAPRPTEAP 169
DB  180 SEVTONTVNE-TWANKKPTNEDMKFPAVPTSHSTMPLNK--TPPATSNIAIMWP----- 229

```

QY	190	FATSSSGVMGRYRPGAINPVRREGTATVAGSTYTSNGMWRSGRBDGLINSMNACTVIG	243
DB	230	-----INCKTIIQGFSSADGC-----NKGIDISGRGAQVAAAAAGRVVY	268
QY	250	ADHNMDGAS--IVIQHTNGFVSSYYTHIDACVKTGDTVTRTGORIAISMKNQPGAALEFR	307
DB	269	AGDALRGVGNLIITIKHNDSYLSAVAHNMSILVKKQGEVKAQGIAMKSSSGNTITKLHRE	328
QY	308	ISRNGVTVDEPLTVL	321
DB	329	IRYKQGSYDPMRKYL	342
RESULT 4			
AA74313	ID	AA74313 standard; Protein; 414 AA.	
XX	XX	AA74313;	
DT	21-MAR-2000	(first entry)	
DE	Neisseria meningitidis ORF 025	protein sequence SEQ ID NO:112-1.	
XX	XX		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;		
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;		
KW	antibacterial; gene therapy.		
XX	Neisseria meningitidis.		
OS			
PN	W0957280-A2.		
PD	11-NOV-1999.		
PF	30-APR-1999;	99WO-US09346.	
XX	01-MAY-1998;	98US-0083758.	
PR	31-JUL-1998;	98US-0094869.	
PR	02-SEP-1998;	98US-0098994.	
PR	02-SEP-1998;	98US-0099062.	
PR	09-OCT-1998;	98US-0103749.	
PR	09-OCT-1998;	98US-0103794.	
PR	09-OCT-1998;	98US-0103796.	
PR	25-FEB-1999;	99US-0121528.	
XX	(CHIR) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;		
PI	Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;		
PI	Teteltn H, Venter JC;		
XX	WPI; 2000-062150/05.		
DR	N-PSDB; AA53075.		
PT	Novel Neisserial polypeptides predicted to be useful antigens for		
PT	vaccines and diagnostics		
XX			
PS	Claim 2; Page 207; 1453pp; English.		
XX	AA553015 to AA254536, AA254577 to AA254615, and AA774253 to AA775941		
CC	represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides		
CC	and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent		
CC	PCR primers used in the exemplification of the present invention. The		
CC	polypeptides, the polynucleotides, antibodies and compositions of		
CC	the invention can be used as vaccines, as diagnostic reagents, and as		
CC	immunogenic compositions. The polypeptides can be used in the		
CC	manufacture of medicaments for treating or preventing infection due to		
CC	Neisserial bacteria (e.g. meningitis and septicemia), to detect the		
CC	presence of Neisseria bacteria, or to raise antibodies. They may also		
CC	be used to screen for agonists or antagonists, which may themselves		
CC	have use as antibacterial agents. The polynucleotides of the invention		
CC	may also be used in gene therapy protocols.		

XX	Sequence	414 AA;	12.2%; Score 203; DB 21; Length 414;
XX	Query Match		Best Local Similarity 27.3%; Pred. No. 5,8e-09;
XX	Matches	71; Conservative 42; Mismatches 97; Indels 50; Gaps 11	
QY	73 YQVKGQDTYKSAQRYGLWREIGHINLNSSYITLGGOMLTLSGDLKYKREISSGVN		132
DB	184 HTIVRQDTYVNIKSRKHISQDDFRANMKTQN-ILISIQ-----IVKKRKGVA---		232
QY	133 TATTPRVAVQSSRRPVQOQHRAVQKPTTPRVVYKKPTTPRVQOQRPVAP-----		183
DB	233 ----PKAAAK-SRPVAP--AAVQTPV-----KPAAPRVQSAFQPAAPAAENKAVA		278
QY	184 PVTEDAPPA---TGSGGVNQFRYPVGAATNPVRRFGATATVAGSTVTSNGMIFSGRDDGLI		239
DB	279 PAQSPVAPASGSRGSRVGIWQRP--TGKRVADFG-----NNKGVADIAGNAGQPV		328
QY	240 NASNACTVIOADHNMDS--IVIQHTNGFVSSYTHIKDAQVKTGDTPTVTRGRIASMKQ		297
DB	329 LAADGIVVYVAGSLGNGYLVLIYIHNSSFLLAYGHQKLLVBEQDQVKKQGVQVALLMNT		388
QY	298 PSGALFEPRISRNQVYVDP 317		
DB	389 EASRTQLHEVRQNGKRVNP 408		
RESULT 5			
XX	AA74311		
XX	AA74311 standard; Protein; 337 AA.		
XX	AA74311;		
DT	21-MAR-2000 (first entry)		
DE	Neisseria gonorrhoeae ORF 025 protein sequence SEQ ID NO.118.		
XX	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;		
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;		
KV	antibacterial; gene therapy.		
XX	Neisseria gonorrhoeae.		
OS	Neisseria gonorrhoeae.		
FN	W09957280-A2.		
PD	11-NOV-1999.		
XX	30-APR-1999; 99MO-US09346.		
XX	01-MAY-1998; 98US-0083758.		
PR	31-JUL-1998; 98US-0094869.		
PR	02-SEP-1998; 98US-0098994.		
PR	02-SEP-1998; 98US-0098994.		
PR	09-OCT-1998; 98US-0103749.		
PR	09-OCT-1998; 98US-0103749.		
PR	09-OCT-1998; 98US-0103796.		
PR	25-FEB-1999; 99US-0121528.		
XX	(CHIR) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
XX	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;		
PI	Petersen J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;		
PI	Tettelin H, Venter UC.		
XX	WPI: 2000-062150/05.		
DR	N-PSDB; AA253075.		
XX	Novel Neisserial polypeptides predicted to be useful antigens for		
XX	vaccines and diagnostics		
XX	Claim 2; Page 206 - 1453pp; English.		

XX	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AA74253 to AA75941
CC	represent novel <i>Neisseria meningitidis</i> and <i>N. gonorrhoeae</i> polymucloectides
CC	and polypeptides. AAZ54537 to AAZ54576 and N. AZ54616 to AAZ5473 represent
CC	PCR primers used in the exemplification of the present invention. The
CC	polypeptides, the polymucloectides, antibodies and compositions of
CC	the invention can be used as vaccines, as diagnostic reagents, and as
CC	immunogenic compositions. The polypeptides can be used in the
CC	manufacture of medicaments for treating or preventing infection due to
CC	<i>Neisseria bacteria</i> (e.g. meningitis and septicemia), to detect the
CC	presence of <i>Neisseria bacteria</i> , or to raise antibodies. They may also
CC	be used to screen for agonists or antagonists, which may themselves
CC	have use as antibacterial agents. The polymucloectides of the invention
CC	may also be used in gene therapy protocols.
SQ	Sequence 337 AA;
Query Match	12.0%; Score 200; DB 21; Length 337;
Best Local Similarity	26.2%; Pred. No. 8e-09;
Matches	66; Conservative 45; Mismatches 109; Indels 32; Gaps 9
OY	73 YVKGQDGTYSKIAQRVGLNWRREIGHINNLSYTYTGQWLTWSGDLKYRERSISSGVN 132 : : : : : : : : : : : : : : : : : : : DB 105 HTIVRDGTYYNISKEHYHISODDFRAMGMTDN-TLTSIGQ-----IVKKPFGVAAPKT 156
OY	133 TATTPRPVVAVSSRPVQQHPAVOKETFPVVVVVKKPPTPPVVQ--PAVPAPVTEA- 188 : : : : : : : : : : : : : : : : : : : DB 157 AAVERSPAPAAPAAQTPEVK--PAAGPP-----VGSAPOPAPAENKNAVPPAPAPPOSAPA 209
OY	189 -PATSGSGVMOPFRYPVGAITNPVVARFGETATVAGSTVSNGMFMFSGRGDLLINASNAGTV 247 DB 210 SPFGRTSVGGIWMQBP--TQGKVADPFQCG-----NRGVDIAAGAQPVLAADGRV 259
OY	248 IODNHMPDAS--IYIOHTNGVSSSYTHIKDAQYKTDPTYRTGORLASMGNOQPSGAALPE 305 DB 260 VVASGLRLRYGNLVIIQHNSSETLATYGHNQCLVLVBEGQYKRQGVALLMGMNTASTQTOH 319
OY	306 FRISRNGCVYPD 317 : : : : : : : : : : : : : : : : : : :
DB	320 FEVRONGKEVNP 331 : : : : : : : : : : : : : : : : : : :
RESULT 6	
AA74312	
ID	AA74312 standard; Protein; 351 AA.
XX	
AC	AA74312;
DT	21-MAR-2000 (first entry)
XX	
DE	<i>Neisseria meningitidis</i> ORF 025 protein sequence SEQ ID NO.110-1.
XX	
KW	<i>Neisseria meningitidis</i> ; <i>Neisseria gonorrhoeae</i> ; antigen; vaccine; antigenic; diagnosis; immunogenic; infection; meningitis; septicemia; antibacterial; gene therapy.
OS	<i>Neisseria meningitidis</i> .
PX	
MO957280-AA2.	
PD	11-NOV-1999.
PR	
30-APR-1999;	99WC-USO9346.
PR	
01-MAY-1998;	98US-0083759.
31-JUN-1998;	98US-0084629.
02-SEP-1998;	98US-0088954.
02-SEP-1998;	98US-0093962.
09-OCT-1998;	98US-0103749.
09-OCT-1998;	98US-0103794.
09-OCT-1998;	98US-0103796.
25-FEB-1999;	99US-0121528.

PA	(CHIR) CHIRON CORP.
PA	(GENO-) INST GENOMIC RES.
PB	Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI	Petersen J, Piza M, Rappoli R, Ratti G, Scalato E, Scarselli M;
PI	Tettelin H, Venter JC;
PR	WPT; 2000-062150/05.
DR	N-PsDB; AA253074.
XX	
PT	Novel Neisserial polypeptides predicted to be useful antigens for
PT	vaccines and diagnostics -
XX	
PS	Claim 2, Page 206; 1453pp; English.
XX	
CC	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC	represent novel <i>Neisseria meningitidis</i> and <i>N. gonorrhoeae</i> polynucleotides
CC	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC	PCR primers used in the exemplification of the present invention. The
CC	polypeptides, the polynucleotides, antibodies and compositions of
CC	the invention can be used as vaccines, as diagnostic reagents, and as
CC	immunogenic compositions. The polypeptides can be used in the
CC	manufacture of medicaments for treating or preventing infection due to
CC	Neisserial bacteria (e.g. meningitis and septicaemia) to detect the
CC	presence of <i>Neisseria</i> bacteria, or to raise antibodies. They may also
CC	be used to screen for agonists or antagonists, which may themselves
CC	have use as antibacterial agents. The polynucleotides of the invention
CC	may also be used in gene therapy protocols.
XX	
SQ	Sequence 351 AA;
Query Match	11.9%; Score 198; DB 21; Length 351;
Best Local Similarity	27.2%; Pred. No.1.3e-08;
Matches	72; Conservative 38; Mismatches 95; Indels 60; Gaps 12,
OY	73 YOVKGGDPVSKAQRVYGILNWRKGHINNLSSTYYITGQMFLTWSGLDKVRERSISSG--130 ::: ::: :: :
Db	121 HTIVAGDVTYNISKRY-----HISODD-----FRAMNG---MTDNMLSIQOI 159 ::: ::: :: :
OY	131 --VNFAHTPPSP-VAAOSSRPVVQAOKFPFPPVVVKKPFTPPVVOQPAPVPAP----183 ::: ::: :: :
Db	160 VKVRKAGTAAPKRTALAVESRPAVP-AAVQTLP-----KRAAOFPVGSAPOPAAPAEN 210 ::: ::: :: :
OY	184 ----PVTEAPPAP----TGSSSGWOFRYPVGAITNPVRRFGATYVAGSITVSNGMFFSGR 234 ::: ::: :: :
Db	211 KAVPAAPAQSPASPSGTRSVGVGWGRP-TQGKVADDFG-----NNKGVDIAGN 260 ::: ::: :: :
OY	235 DGDLLINASNAGTVIQADHRMDAS--IYIGHNMGVSSIIHKDAQYKTGTDTVFPGORIA 292 ::: ::: :: :
Db	261 AGOPVLTAADGKVVAVAGSLRGYGNLVIIIOHNSFLTAYGHNOKLVBEGDOQVKGOOVA 320 ::: ::: :: :
OY	293 SMKNOPEGALLPEFRISRNGVYNDP 317 :::
Db	321 LMGNLTDASTRQLHFVEYRONKEVNP 345 :::
RESULT 7	
ID	ABB48721 standard; Protein; 436 AA.
XX	
AC	ABB48721;
XX	
DT	05-FEB-2002 (first entry)
XX	
DE	Listeria monocytogenes protein #1425.
XX	
KW	Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW	Vitamin B12; bacterial infection; disease.
OS	Listeria monocytogenes.
XX	
XX	WO200177335-A2.
XX	

```

PD 18-OCT-2001.
XX
XX 11-APR-2001; 2001WO-FR01118.
XX
XX 11-APR-2000; 2000FR-0004629.
XX
XX (INSP ) INST PASTEUR.
XX
XX Buchrieser C, Frangeul L, Couve E, Ruanlok C, Fajhi H, Dehoux P,
XX Dussauget O, Chetoui F, Nedjat H, Glaser P, Kunst F, Cossart P,
XX Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Bojand JA,
XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
XX Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,
XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
XX Madeno E, De Pablos B, Wehlant U, Kaerst U, Entian K, Hauf J,
XX Rose M, Voss H;
XX
XX WPI, 2002-010914/01.
XX
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides -
XX
XX Claim 6; SEQ ID No 1426; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and
XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX B12. The genome sequence and proteins encoded by it are also useful for
XX selecting compounds that regulate gene expression and cell replication
XX and modulate L. monocytogenes-related diseases. In addition, the genome
XX sequence and proteins encoded by it are useful in pharmaceutical and
XX vaccines compositions for the treatment or prevention of infections by L.
XX monocytogenes and related organisms.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 436 AA;
XX
XX Query Match 9.5%; Score 158.5; DB 23; Length 436;
XX Best Local Similarity 24.8%; Pred. No: 3.8e-05;
XX Matches 66; Conservative 33; Mismatches 100; Indels 67; Gaps 11
XX
XX 76 KQGDVSKIAQRYGLNREIGHINNLSYYITGOWLTMSGDLYKVERSISSGVNPAH 135
XX :::::
XX 210 EKNDLVMLANKKUL-----KSEQLINSE-----QGLTDEKRLASIAIEK 254
XX
XX 136 TPSPVAVOSSRPVQGHAVOKPTPPVVVVYKKPTPPPVVQOPAVAPVTEAPATGSS 195
XX :::::
XX 255 AKOEALIKAAIEKMQEAAASAASAKSAVVK-----QPSSSSEATE---TVSS 300
XX
XX 196 GVMQFRYVGA-----TNPVVRP-----GTATVAGTNTVNSGMFFSRDGLINAS 242
XX :::::
XX 301 GGGQPIKASGILTSGFSERINPTVTKESHKGDIAAGGTVT-----VSPA 347
XX
XX 243 NAGTVIQADHNMDS-----IVIQHTNGFVSYSIHKDAQV--TGDVTVTGORTAS 293
XX :::::
XX 348 ASGTIVFSGFASGSGFGCYGVVKKIDIGNGFQTLTGMRAGSLKVTYGGQVSSQGPIGI 407
XX
XX 294 M-KNQPSCALFEFRISRNQVYDP 317
XX :::::
XX Db MGSTGQSTGQHL-HFEIKHNGIPIVD 432

```

ID AAR45178 standard; Protein; 478 AA.
 AC AAR45178;
 XX
 DT 16-JUN-1994 (first entry)
 XX
 DE Listeria p60 protein.
 XX
 KW Immunogenic polypeptide; antibodies; immunoassay; conjugate.
 OS Listeria monocytogenes.
 XX
 PN DE4318450-A.
 XX
 PD 16-DEC-1993.
 XX
 PP 03-JUN-1993; 93DE-4318450.
 XX
 PR 11-JUN-1992; 92DE-4219111.
 XX
 PR 23-NOV-1992; 92DE-4239567.
 XX
 PA (MERB) MERCK PATENT GMBH.
 XX
 PI Schubert P, Neumann S, Pawelzik M, Linxweiler W, Burger C,
 XX
 PI Hofmann G, Hubert A, Goebel W, Koehler S,
 XX
 DR WPI; 1993-406956/51.
 XX
 PT New primers for PCR detection of Listeria - including individual
 PT species, also new peptide(s) for raising antibodies for
 XX
 XX immunochromatological detection
 PS Disclosure; Fig 4; 19pp; German.
 CC The sequence is that of the Listeria p60 protein. Antibodies
 CC generated against the protein can be used in the detection
 CC of Listeria by immunoassay (partic. ELISA). The detection method
 CC allows determination of individual Listeria species, esp.
 CC L. monocytogenes.
 SO Sequence 478 AA;
 Query Match 9.1%; Score 151.5; DB 14; Length 478;
 Best Local Similarity 21.3%; Pred. No. 0.00017; Mismatches 117; Indels 143; Gaps 15;
 Matches 83; Conservative 46; Mismatches 117; Indels 143; Gaps 15;
 QY 25 ITTCLAGCASKPTYNSTSGSGSHRTSGSGGLA--IGSQVTD----- 65
 DB 102 IITSIKGG--TKVIVETTESNGWHKITYNDGKTGFVNGKYLTDKAVSTPVAFTQEVKKT 159
 QY 66 --SQGVP-----NRVQVKGADPTVSKIAQRYG 89
 DB 160 TTOQAAVAETKTEVKQTQATTPARKVAETETPIDONATTAAVKSGLTIALSVKYG 219
 QY 90 LMRREIGHINNLSSTIYTGOMLTMSGDLKVRERSISSGVNTAHTPSVVAOSRPPV 149
 DB 220 VSVQDIMSNNLSS--SIYVGOKLAI-----KOTANATPKAVKT----- 259
 QY 150 QQHFAVQKTPPVV-----VKKPTPPPVVQAPVAPVTE-----APFATG 193
 DB 260 -EAPAEKQAAFPVENNTNTNTATTEKEKATQ--OQTPAKP--TEAAKPAPAPSTNT 313
 QY 194 SSGVMOFRYPVATPVVRRFGATVAGSTVTSNGMFGSGRDDLINAGTAVIOADHN 253
 DB 314 NANKNTNTNTNTNTNTPSKNTNTNTNTNTNTNSNTANOGSSNNNSNSASALIAEQKH 373
 QY 254 MD-----GASVIOHTNGF--VSSYIHIKDAQVGTGQVTRG 288
 DB 374 LKQASWCGNGPTTDCSGYTKVFAAGISLPRISGAGVASTTRISBQAKPDDLV--- 430
 QY 289 QRIASMKQPSGALLFFP--RISRGVTV 315
 DB 431 -----FFDYSGSISHVGIIV 445

RESULT 9
 ID ABB53578 standard; Protein; 439 AA.
 AC ABB53578;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein acma.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 OS Lactococcus lactis IL1403.
 XX
 PN FR2807446-A1.
 XX
 PD 12-OCT-2001.
 XX
 PP 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX
 PI Bojorine A, Sorokine A, Renault P, Ehrlich SD,
 XX
 DR WPI; 2002-043418/06.
 XX
 PT New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species -
 XX
 XX Claim 6; SEQ ID No 280; 2504bp; French.
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of the invention are useful for the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO2001/77334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 SO Sequence 439 AA;
 Query Match 8.8%; Score 146.5; DB 23; Length 439;
 Best Local Similarity 23.6%; Pred. No. 0.00041; Mismatches 69; Conservative 39; Mismatches 107; Indels 77; Gaps 10;
 Matches 69; Conservative 39; Mismatches 107; Indels 77; Gaps 10;
 QY 26 TTCLLAGCASKPTYNST-----SSGSHRTSGSGGLAIGSQVITDQVPR 72
 DB 184 TALLGKATDPNTGALNRIISQYNLTRFDGASAGTSGG--STANTNNNSNTSSTT 242
 QY 73 YOVQGGTVSKIAQRYGLMRREIGHINNLSSTIYTGOMLTMSGDLKVRERSISSGVN 132
 DB 243 YTVASGDTLNGISQKIGISVAQIOSANNLKST-VYIGOKLVLTSSSSNTNTSSSSGN 301
 QY 133 TATHPSPVAVOSSRPPVQOHPAVQKPTPPVVVKKPTPPPVVQAPVAPVPT---EA 188
 DB 302 SAGTTTPTT-----SVTPAPASQTIITKSGDTLNGLSV 336
 QY 189 PFATSSGGMORRYPVGATNPVRRFGATVAGSTVTSNGMFGSGRDDLINAGTAVI 248
 DB 337 KYVTITIAQKSMWHLNLSPTIFIGNLTIVSOSAGSSSSSTG-----SSAST-- 382
 QY 249 QADHNNDGASIVIOHTNGFVSSYIHIKDAQVGTGTV-----RTGRIASMK 295
 DB 383 SETSNSSAAS-----NTSIH--KVVKGDTLWGLSGSGSPIASIK 420

RESULT 10
 ABB48659 standard; Protein; 482 AA.
 ID ABB48659 standard; Protein; 482 AA.
 AC ABB48659,
 DT 05-FEB-2002 (first entry)
 DE Listeria monocytogenes protein #1363.
 XX Listeria monocytogenes protein #1363.
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KM vitamin B12; bacterial infection; disease.
 XX Listeria monocytogenes.
 OS WO200177335-A2.
 XX 18-OCT-2001.
 PD 11-APR-2001; 2001WO-FR01118.
 PF 11-APR-2000; 2000FR-0004629.
 PR (INSP) INST PASTEUR.
 PA Buchrieser C, Frangoul L, Couve E, Ruenick C, Fathi H, Dehoux P,
 PI Duesurget O, Chetoui F, Nedjati H, Glaeser P, Kunst F, Coesart P,
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
 PI Chakraborty T, Doman E, Hain T, Berche P, Charblat A, Durand L,
 PI Perez-Diaz J, Baguerio F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno B, De Pablo B, Weiland J, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H,
 XX WPI; 2002-010914/01.
 DR Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides
 XX Claim 6; SEQ ID No 1364; 192pp; French.
 PS The present invention relates to the genome sequence of Listeria
 XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIDO
 CC at ftp.wido.int/pub/published_pct_sequences.
 XX
 XX Sequence 482 AA;
 SQ
 Query Match 8.8%; Score 146.5; DB 23; Length 482;
 Best Local Similarity 20.6%; Pred. No. 0.00046;
 Matches 81; Conservative 46; Mismatches 121; Indels 145; Gaps 14;
 QY 25 ITTCLACASKPTNSTSGSGSHRTSGSGLA--IGSOVITD----- 65
 DB 100 IITSTKGG--TKVIVETTESNGMHTKITYNDGTGTVNGSKYLDKAVSPVPAOTGVKKT 157
 QY 66 --SGGVP-----NRGVKOGDVTWSKIARNG 89

DB 158 TTQQAAPAAETKTEVKQTQATTPAPKVAETKETPVVDQNAATHAVSGDTIMALSVKYG 217
 QY 90 LNMREIGHINNLNSYTYTYGQWLTLMSGDLKVAERISISSGVNTAHPSPVAVOSSRPV 149
 DB 218 VSVQDIMSWMNLSSS-SLYVGQKLA-----KQANATATPAEVKT----- 257
 QY 150 QQHFAVOKETPPVV-----VKKPPTPPVVOAPVAPVTEAPATGSSGVWQ 199
 DB 258 -EAPPAEKQAPVVKENTNTNTATTEKETITQ---QOTAPKATPAKPAARSTVINA 313
 QY 200 FRYPVGA-----TNPVRRFGTATVAGSTVTSNGMFGSRDGLINASNAGTVIQ 249
 DB 314 NKTNTNTNTNTNTNTNTNTPTPSKNTNTNTNTNTNTNTNANQSSNNNSASAIIAE 373
 QY 250 ADHMD-----GASVIGHITNGF-VSSYHIHMDAQVTKGDT 284
 DB 374 AQKHGKAVYSGNGPPTFDGSGYTKYFAPAKAGISLPTSGAQYASTTRISEQAKPGDL 433
 QY 285 VRTGORTASMKNOPSGAALPEF--RISNGVYV 315
 DB 434 V-----FPDYGSGISVIGIYV 449

RESULT 11
 AAR73913
 ID AAR73913 standard; protein; 484 AA.
 AC AAR73913;
 DT 05-DEC-1995 (first entry)
 DE Listeria monocytogenes protein p60 precursor.
 XX Listeria monocytogenes, protein p60 precursor; vaccine;
 KM meningitis related homologous antigenic sequence; MRHAS; RV-1;
 KM immunosassay; diagnosis; treatment; prophylactic; bacterial;
 XX viral.
 OS Listeria monocytogenes.
 XX WO9509232-A.
 EN 06-APR-1995.
 PD 28-SEP-1994; 94WO-CA00516.
 PF 28-SEP-1993; 93US-0127499.
 PR (SHAR/) SHARMA L R.
 PA (VALS/) VAN ALSTYNE D.
 PI Sharma LR, Van Alstyne D;
 DR WPI; 1995-147431/19.
 PT New peptide(s) and corresp. antibodies for the treatment of
 PT meningitis - the peptide(s) corresp. to homologous antigenic
 PT sites on bacterial and viral agents and on chemokine(s), used for
 PT detecting and preventing meningitis
 XX
 XX Claim 47; Fig 7/10; 98pp; English.
 CC AAR73913 is the Listeria monocytogenes protein p60 precursor. It
 CC contains the meningitis related antigenic sequences (MRHAS) claimed
 CC in AAR73891-R73894 and AAR73903-R73906, which are recognised by a
 CC monoclonal antibody from the hybridoma Rubella virus (RV)-1. The
 CC claimed MRHAS peptides may be used in immunoassays to diagnose the
 CC presence of bacterial and/or viral meningitis agents in a sample,
 CC or in prophylactic and therapeutic meningitis treatments. The
 CC peptides may also be used as vaccines against meningitis.
 CC NB: Identified by matching corresponding MRHAS peptides.

```

PT useful as components of vaccines for treating bacterial infection such
PR as otitis media, delayed speech learning and inflammation of middle ear
PS _
SS
XX Claim 3; Page 88; 90pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC several versions of the BASB201 protein from non-typeable Haemophilus
CC influenzae. These can be used in the production of vaccines against H.
CC influenzae infection, which can cause otitis media in infants and
CC children, pneumonia in elderly sinusitis, nosocomial infections, or
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infections of the upper respiratory tract and inflammation of
CC the middle ear. The present sequence is a version of the BASB201 protein
CC sequence of the invention.
CX
SQ Sequence 410 AA;

Query Match
Best Local Similarity 8.0%; Score 134; DB 23; Length 410;
Matches 62; Conservative 37; Mismatches 111; Indels 42; Gaps 10

QY 74 QVVGQDTVVKIKORGLMWREIGHINNLSYTIYTQWLTLMWGDKYRERSISSGVNT 133
Db 188 OKKQGQALOKAQKEHQFS-----LNLNINMLDLDPK-----LNALKNEGALNQEIQR 236
QY 134 AHTPSPAVAVQSSRPVV-QQHPPAVQKPTPPPVVVAKKPTTTPPVVQQPAPVPYTEAPART 192
Db 237 AQGARBEKKERELALROQAEBEKT-----SKTYQTVDBROLN 278
QY 193 GSSGV---MFRYPVGNATNPVVRFGTAIVAGSTVTSGNMFGSRDDGLINAASNACTVI 248
Db 279 STSGGLAAKQOYSLEPVSQS--ILHTPG--SIQAGEVWKMGVIGASACTPYKALIAAGRVI 334
QY 249 QAHD-NMDGASIVIQHTNGFVSSYIHIDQAVKGTGDTVRTGORIASMKQ--QPBGAALEF 305
Db 335 LAGVLNIGVMVIVKKGETDLSLYGPNQAVSVKQLVASQGVIAQVONTGEISRSALY- 393
QY 306 FRISRNCNVYPD 317
Db 394 FGISRKGTFPVN 405

RESULT 13
AA017656
ID AA017656 standard; Protein; 410 AA.
AC
XX AA017656;
XX
DT 05-AUG-2002 (first entry)
DE H influenzae BASB201 #1.
KM BASB201, otitis media; pneumonia; sinusitis; nosocomial infection;
KW auditive nerve damage; delayed speech learning; vaccine;
XX antibacterial; auditory; antiinflammatory.
OS Haemophilus influenzae.
XX
FH Key
FT Location/Qualifiers
FT Domain 32..300
FT /label= KEQ_rich_domain
FT Domain 301..400
FT /label= peptidase_M37-like_domain
PM WO200230967-A2.
XX
PD 18-APR-2002.
XX
PE 05-OCT-2001; 2001WO-EPI1561.
PS

```

PT	useful as components of vaccines for treating bacterial infection such
PR	as otitis media, delayed speech learning and inflammation of middle ear
XX	-
XX	Claim 3; Page 88; 90pp; English.
PS	The present invention provides the protein and coding sequences of
CC	several versions of the BASB201 protein from non-typable Haemophilus
CC	Influenzae. These can be used in the production of vaccines against H.
CC	children, pneumonia in elders, sinusitis, nosocomial infections, or
CC	accumulative diseases, chronic otitis media with hearing loss, fluid
CC	accumulation in the middle ear and/or auditory nerve damage, delayed speech
CC	learning, infections of the upper respiratory tract and inflammation of
CC	the middle ear. The present sequence is a version of the BASB201 protein
CC	sequence of the invention.
XX	
SQ	Sequence 410 AA;
Query Match	8.0%; Score 134; DB 23; Length 410;
Best Local Similarity	24.6%; Pred No. 0.0043;
Matches 62; Conservative 37; Mismatches 11; Indels 42; Gaps 10;	
QY	74 QVKGQGTVAISKIAQRGLMREIGHINLNSYYITYGQMTLWSGDLKVERSSISGVT 133
Dy	188 OKKOQOALOKAQOEHQFS-----LNLLNQLMLDDDK-----LNALKAMEGALRQEIOR 236
OY	134 AHTPSPVAVOSSRPV--QQHPAVKKPFPVVVVVKFPPRPVVOQPAPVPAPVTEAPPAT 192
Dy	237 AEQAAREQEKERRELAQRKAEEKT-----SKFYOPTVEROLLN 278
OY	193 GSSGV----MOFRYPVATNPVVRREGCTAVASGYTSNGMFSGRDDLLINANAGTVI 248
Dy	279 STSGIGAAKQOYSLPVGS--LIHTG--SIQAGEKRWKAWTGASACTPVKAIATAARVI 334
OY	249 QAHD-NMDGASIVYOHNTGNFVSXYIHAKDAQYKTGDIVRTGORIASMKN--OPEGALTFE 305
Dy	335 LAGLYNGYGMVIVKGETDLSLXGPNOVASVKQGLVASQVIAOVGNNGEISRSLTY- 393
OY	306 FRIRSNQVYDP 317
Dy	394 FGISRKOTPVNP 405
DB	
RESULT 13	
ID	AAOI7656
TD	AAOI7656 standard; Protein; 410 AA.
XX	
XX	AAOI7656;
XX	
XX	05-AUG-2002 (first entry)
XX	H influenzae BASB201 #1.
XX	
XX	BASB201; otitis media; pneumonia; sinusitis; nosocomial infection;
KM	auditive nerve damage; delayed speech learning; vaccine;
KW	antibacterial; auditory; antiinflammatory.
XX	
OS	Haemophilus influenzae.
XX	
FT	Key
FM	Location/Qualifiers
FT	Domain
FT	/label= KEQ_rich_domain
FT	301..400
FT	/label= peptidase_M37-like_domain
XX	
PM	WO200230967-A2.
XX	
PX	18-APR-2002.
XX	
PX	05-OCT-2001; 2001WO-EPI1561.
XX	
XX	13-OCT-2000; 2000GB-0025169.
DR	

PF 29-JAN-2001; 2001MO-CA00128.
 XX
 PR 27-JAN-2000; 2000US-0177711.
 XX
 PA (ECOP-) ECOPIA BIOSCIENCES INC.
 XX (FARNET) FARNET C.
 PI
 PI Staffa A, Zazopoulos E, Mercure S, Nowacki P;
 DR WPI; 2001-476185/51.
 XX N-PSDB; ABL50562.
 PT Novel isolated gene cluster encoding polypeptides involved in
 PT evernomicin biosynthesis useful for construction of evernomicin
 PT overproducing strains, and to allow chemical modifications of
 XX evernomicin to enhance certain properties
 PS Claim 15; Fig 1; 181pp; English.
 CC ABL50555 to ABL50562 represent contigs 1 to 8 from the Micromonospora
 CC carbonacea evernomicin biosynthetic locus gene cluster. The contigs
 CC 49, given in ABL50561 to ABL50590. The gene cluster is useful for the
 CC construction of the evernomicin antibiotic in overproducing strains,
 CC and to allow chemical modifications of evernomicin to enhance certain
 CC properties via genetic manipulations or combinatorial biosynthesis. The
 CC gene cluster can be used to produce genetic systems and genes encoding
 CC novel enzyme activities and avoid the problems of low yield and quality
 CC of evernomicin produced by chemical synthesis.
 SO Sequence 518 AA;
 Query Match 7.8%; Score 130.5; DB 22; Length 518;
 Best Local Similarity 27.0%; Pred. No. 0.012; Mismatches 112; Indels 53; Gaps 14;
 Matches 69; Conservative 22;
 DB 20 LIFGVTTCILAGASPTNSTSGS-----SHRTSG-SGGAL- -GQV- - - 63
 QY 8 LAVGTATLVAAGLGLTPASMAATGCVAVYVOOSGTFGSGNVAITNLSALAGMTL 67
 DB 64 ----TDSQGVNPRYGV- -KODTVSKIAQRYGLNWRREIGHINNINSYTYITQWLT 115
 QY 68 TDFPTSGQOVTOGMEATWSOGSISVMAAS- -LSWN- -GSLGTGSGTTCFNS- - - 117
 DB 116 WSGDLKVRERSISSGVN-TAHTPSPVAVOSSRPVVOOHDAVOKPTPPVVKKPTPPV 174
 QY 118 WSGSNPVPKSFALNCTGCTGTSVTPPTPTTP- - -PTTPTPTPTPTPTPTPTPT 168
 DB 175 VQOPAPVAPVTEAPFATGSGVWQPRYPVGTATNPVRRFGTAIVAG- -STVTSNGWFS 232
 QY 169 T--PPTTPPTTGAAPALKVSG- - -NRLVLTASGATYRLIGVNPASGEFACVGGKGMDS 222
 DB 233 GR-DGDLINASNAGTV 247
 QY 223 GPVDOASVNMKAMNI 238
 DB
 RESULT 16
 ID ABB54367 standard; Protein; 894 AA.
 XX ABB54367;
 AC
 XX 16-MAY-2002 (first entry)
 DT
 XX Lactococcus lactis protein p1244.
 DE
 XX Lactococcus lactis protein p1244.
 KM
 XX Biosynthesis, biodegradation, lactic bacterium, yogurt, cheese.
 OS Lactococcus lactis IL1403.
 XX
 XX FR2807446-A1.

PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMICQUE.
 XX
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 DR WPI; 2002-043418/06.
 XX New nucleotide sequence useful in the identification of Lactococcus
 XX lactis and related species
 PS Claim 6; SEQ ID No 1069; 2504pp; French.
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (AB290521) and related proteins (AB290521-AB290521). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC related species. The proteins, particularly the proteins of the invention
 CC biosynthesis or biodegradation of the invention are useful for the
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 SO Sequence 894 AA;
 Query Match 7.8%; Score 130.5; DB 23; Length 894;
 Best Local Similarity 23.6%; Pred. No. 0.023; Mismatches 112; Indels 79; Gaps 14;
 Matches 69; Conservative 32;
 DB 29 ILAAGCAKPTVNTSSGSHRTSGSGGLAIGSOVITPDSQVPR- -YQK- - - -Q 77
 QY 620 LLSAGCEBDANPTDDE- - -GRRPGRGVWQWMTDSSGASGRVYMINMTAGVTDN 675
 DB 78 GDTVSKIAQRYGLNWRREIGHINNINSYTYITQWLTWSGDLKVRERSISSGVTHPT 137
 QY 676 PDIIT--AQFKLW- - -HQN- - -GOWTAKSSYPSWTOPTMLTININ- - - 717
 DB 138 SPVAVOSSRPVVOOHDA- - -YOKPTPPVVKKPTPTPPVVOOPAPVAPVTEAPATG 193
 QY 718 TQAFVANFERDLNCHPERSTWAGSWYKFKVINKIPSGGGYI- - -APISPIIT- - -VTS 770
 DB 194 SSGVWQFRYPVGTATNPVRRFGTAIVAGSTVTSNGWFSGRDGLIN- - -ASNAG 245
 QY 771 EMG- - -WTSPTTGAQEFNAM- - -DLVNGNPTPTPLASDGG 806
 DB 246 TVIQADHN- - -MDGASIVIOHTNGFVSYYIHKDAQYKGTDTVRTGORTASM 294
 QY 807 QVVOAGSNYYDWYGVYTVIKHADGLYGVYHQSRIQVSVGVQVKKGGQIGLM 858
 DB
 RESULT 17
 ID AAO17657 standard; Protein; 410 AA.
 XX AAO17657;
 AC
 XX 05-AUG-2002 (first entry)
 DT
 XX H influenzae BASB201 #2.
 DE
 XX BASB201; otitis media; pneumonia; sinusitis; nosocomial infection;
 KM auditive nerve damage; delayed speech learning; vaccine;
 OS antibacterial; auditory; antiinflammatory.
 XX Haemophilus influenzae.
 XX
 XX Key Location/Qualifiers

Best Local Similarity 22.7%, Pred. No. 0.11;
Matches 80; Conservative 37; Mismatches 120; Indels 116; Gaps 19;

QY 27 TCILAGC---ASKRTVSTSGSGHRTSGSGGLAI-----GSGVITTSQGVPRRYQVK 76
Db 574 ITVVDGAHLEANGPEEYVLSPDASRQSMGAGSHSLTYELTPAGLQVKISSNGLDCTATFP 633
QY 77 QGDVTS-----KIAQRYSLN--WR-----EIGHI---NN 100
Db 634 XGAPSAAPGEVXAFCSALYRYNFTQRHSLTGMLMHPGLGTFPPSPGHIMESANP 693
QY 101 INSSYTYTGOMTLI--NSGDLKVERSSISSGVNTAHTPSPFAVQSSRPV----- 149
Db 694 FCGEGTLVTRTWSTSGFSSDPSPPEAAPASAAAPGLPYPT-----PVSIDIWLPSPS 747
QY 150 -QQH-----PAYQKP---TPPVVYVKKPTPTPPVVOQAPAPVAPVTEAPFATSSGVNQ 199
Db 748 EESHVDASVSPSEPPAGLTSPIVLT--PPPPPPVAKRATSPFPATRR-----IL 796
QY 200 FRYPVGATNPVVRFRGTATVAGSTVTSNGMFPSSGRDGLIMASNACTVIQADHNMGASI 259
Db 797 YTPDGA-----KYVAGSLXESDCDW-----LVNASNPQ-----HRQGG-- 831
QY 260 VIOHTNGFVSSYIHKDAQVKTGDTVRTGORIASMKNP--SGAALFEFRISHN 311
Db 832 -LCH-----AFYQRFPEAFYSTERIMEGLAAYLTLPRIHADVADYHVEQN 878

RESULT 19
AAR24147
ID AAR24147 standard; Protein: 374 AA.
AC AAR24147;
XX 15-NOV-1992 (first entry)
DT 15-NOV-1992 (first entry)
XX Beta-lytic protease.
DE Beta-lytic protease.
XX Gram-negative; Gram-positive; bacteria; decomposition.
OS Achromobacter lyticus.
XX JP04108387-A.
PN 09-APR-1992.
PD 09-APR-1992.
XX 29-AUG-1990; 90JP-0225136.
PF 29-AUG-1990; 90JP-0225136.
XX 29-AUG-1990; 90JP-0225136.
PR 29-AUG-1990; 90JP-0225136.
XX (WAKP) WAKO PURE CHEM IND.
PA (WAKP) WAKO PURE CHEM IND.
XX WPI; 1992-171653/21.
DR N-PSDB; AAQ25083.
XX The protein sequence of the beta lytic peptidase from Achromobacter
CC lyticus was deduced from the DNA sequence obtd. by PCR using
CC primers based on the sequence of the L. enzymogenes beta-protease.
CC The beta-lytic protease is expected to be an enzyme which can decompose
CC not only Gram-positive bacteria but also some Gram-negative bacteria.
XX

PS Claim 2; Fig 1; 13pp; Japanese.
PT Beta-lytic protease gene and DNA encoding it - for decomposing
PT Gram-positive and some Gram-negative bacteria
XX

Sequence 374 AA;
SQ

Query Match 7.6%; Score 126.5; DB 13; Length 374;
Best Local Similarity 25.9%; Pred. No. 0.017;
Matches 50; Conservative 29; Mismatches 65; Indels 49; Gaps 7;

QY 117 SSDLKVERSSISSGVNTAHTPSPFAVQSSRPVVOQHAPVOKPTPPVVVKKPTPPPVQ 176

Db 154 AGRAAARRRRPAGL-----RPVQRTAPGGGGRGP-----LR 186
QY 177 QAPAPVPTTEAPRATSSGIMQFRVGNATNPVRRFGATYAGS-----YVTSNG 228
Db 187 QGRP-----GRAVSPNGLQFPFPGASWHV--GGAHNTGSGNTPMSSLDMSRGG 236
QY 229 MPFSGRDDLIINASNACTVIQADHNMGASIVIOHTNGFVSSYIHKDAQVKTGDTVRTG 288
Db 237 GWSGNQNNWVWASASAGSFKR--HSCFAEIV--HTGGMSTIYHLMNIQYNTGNAVSMN 292
QY 289 ORIASMKNOPSGA 301
Db 293 TAINPANTQQA 305

RESULT 20
AAV44646
ID AAV44646 standard; Protein: 374 AA.
AC AAV44646;
XX 18-APR-2000 (first entry)
DT 18-APR-2000 (first entry)
XX Achromobacter lyticus beta-lytic protease.
DE Beta-lytic protease; antibiotic; anti-staphylococcal;
XX gene therapy; mastitis; staphylococcal infection; ruminant; cow;
KM transgenic animal; altered beta-lytic protease gene.
XX

OS Achromobacter lyticus.
XX WO967381-A1.
FN 29-DEC-1999.
PD 29-DEC-1999.
XX 22-JUN-1999; 99WO-US14073.
PF 22-JUN-1999; 98US-0090175.
XX 22-JUN-1999; 98US-0090175.
PR 21-JUN-1999; 99US-0337079.
XX (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
PA Bramley JA, Plant KI, Kerr D;
XX WPI; 2000-147208/13.
DR N-PSDB; AA249721.
XX Treatment of Staphylococcal infections, such as mastitis, in ruminant
PT animals, especially cows -
PT
XX Disclosure; Fig 14B; 61pp; English.
PS
XX The present sequence is a Achromobacter lyticus beta-lytic protease
CC which has anti-staphylococcal activity. The beta-lytic protease gene is
CC used in the production of altered genes which
CC allow expression and preferably secretion of active protein in mammalian
CC cells/tissues. The altered gene is produced by operably linking the
CC beta-lytic protease coding sequence with mammalian promoter, signal
CC peptide and translation initiation sequences. The modified sequence
CC is used in gene therapy to treat staphylococcal mastitis infections in
CC ruminants, e.g. goats, sheep, and cows. It is also used to produce
CC transgenic animals which are resistant to staphylococcal infections.
XX

Sequence 374 AA;
SQ

Query Match 7.6%; Score 126.5; DB 21; Length 374;
Best Local Similarity 25.9%; Pred. No. 0.017;
Matches 50; Conservative 29; Mismatches 65; Indels 49; Gaps 7;

QY 117 SDDLKVERSSISSGVNTAHTPSPFAVQSSRPVVOQHAPVOKPTPPVVVKKPTPPPVQ 176
Db 154 AGRAAARRRRPAGL-----RPVQRTAPGGGGRGP-----LR 186

QY 177 QPAPVAPVTEAPFATGSSGWMQFRYPVGATPVVRRFGTATVAGS-----TGTENG 228
 DB 187 QGRP-----GRAVSPNGILQFPFPGASWHV---GGATHTGSSGMYPMSLDMSGG 236
 QY 229 MWFSGRGGDLINASNAGTVIQADHNMDSASIVIOHTNGFVSSYHIKDAQVKTGDTV 288
 DB 237 GWSGNQNONWVWSASASAGSFKR--HSSCFAEIV--HTGWMSTTYHIANIQTNTGANVSMN 292
 QY 289 QRASMKKQPSGA 301
 DB 293 TAIANPANTQAQA 305
 RESULT 21
 AAG92487
 ID AAG92487 standard; Protein; 205 AA.
 AC AAG92487;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 6241.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 EN EPI108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PP 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI: 2001-376931/40.
 DR N-PSDB; AAH67706.
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 17, SEQ ID NO: 6241; 246bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SQ Sequence 205 AA;
 Query Match 7.5%; Score 125.5; DB 22; Length 205;
 Best Local Similarity 33.8%; Pred. No. 0.0093;
 Matches 52; Conservative 14; Mismatches 73; Indels 15; Gaps 5;

QY 179 APVAPVTEAPFATGSSGWMQFRYPVG--ATNPVRRFGTATVAGSTVTS----- 226
 DB 41 ATMPAPASQIDYAGLSSGVADTVAAGAVNTAATVAPATVARNPFTSGFGRMGTFH 100
 QY 227 NGMWFSGRGGDLINASNAGTVIQADHNMDSASIVIOHTNGFVSSYHIKDAQVKTGDTV 285
 DB 101 NGDIANSIGTPIYAVWAGTVISSGPASGCGQMRIQHDGDSISYGHMEYLYVSGERV 160
 QY 286 RTGORIASMKKQP--SGAALPEFRISRNGVY-VDP 317
 DB 161 AAGGEIAGMSQGFSTGSHLFEIHPDGVTPVDP 194
 RESULT 22
 AAM95501
 ID AAM95501 standard; Protein; 288 AA.
 AC AAM95501;
 XX
 DT 26-MAR-1999 (first entry)
 XX
 DE B. subtilis Yoch related gene product YochHsubt.
 XX
 KW RP-factor; resuscitate; latent cell; growth-stimulation; receptor; YABE;
 KW convertase; vaccine; immunotherapy; prophylaxis; excipient; antibiotic;
 KW antimicrobial; tuberculosis; food; pharmaceutical; culture additive.
 XX
 OS Bacillus subtilis.
 XX
 EN WO9855624-A1.
 XX
 PD 10-DEC-1998.
 XX
 PP 03-JUN-1998; 98WO-GB01619.
 XX
 PR 27-MAY-1998; 98GB-0011221.
 PR 04-JUN-1997; 97GB-0011389.
 XX
 PA (UYWA-) UNIV WALES.
 XX
 PI Kaprelyants AS, Kell DB, Mukamolova GV, Young DI;
 PI Young M;
 DR WPI: 1999-070270/06.
 XX
 PT New bacterial resuscitation factors - useful for stimulating latent
 PT bacteria and growth, particularly for use as antimicrobials, also
 PT vaccines comprising bacteria with mutations in resuscitation factor
 PT genes
 XX
 PS Disclosure; Fig 1B; 76pp; English.
 XX
 CC The invention relates to RP-factors which are substances that can
 CC resuscitate dormant, moribund or latent cells, possibly also having
 CC growth-stimulating activity. Host cells containing a vector comprising
 CC the RP-factor or its receptor encoding nucleic acid can be used for the
 CC recombinant production of the RP-factor. RP-factors, their receptors or
 CC convertases, antibodies (against the RP-factors and RP-factor receptors
 CC or convertases), antagonists and agonists, are useful in vaccines and for
 CC immunotherapy, diagnosis and prophylaxis. They are also useful as
 CC excipients, generally as antimicrobials and especially for infections
 CC associated with latency. They can be used as potentiators of antibiotics
 CC such as isoniazid, streptomycin etc., in treatment of tuberculosis.
 CC RP-factor can also be used: to determine the microbiological quality of
 CC foods, pharmaceuticals, medical products; as culture additive for
 CC bacteria; to stimulate growth and/or to resuscitate microorganisms and to
 CC produce libraries of biomolecules and microorganisms (which may then be
 CC screened for useful products). Probes complementary to the RP-factor
 CC nucleic acid are used to identify and clone other RP-factor genes. The
 CC inventions may also be used to prevent bacterial resuscitation. Breaking
 CC dormancy with RP-factor facilitates detection, culture and enumeration of
 CC many bacteria. The present sequence represents a B. subtilis Yoch related

CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX SQ Sequence 639 AA;

Query Match 7.2%; Score 120.5; DB 22; Length 639;
 Best Local Similarity 21.7%; Pred. No. 0.11;

Matches 54; Conservative 32; Mismatches 84; Indels 79; Gaps 7;

QY 24 VITTCILAGCASKPTYNSTSGSHRTSGGGLAIGSGVITDSOG----- 68
 DB 432 VITKRLILSSASSPFIQSSYGVGINQRFHAAHKQSQADDDQASGLQSPPSRLSPTL 491
 QY 69 -----VNNRYQVKQG--DVSKIAQRYGLNWRREIGHNNLNSSYTIYTG 110
 DB 492 IDNSAKQLARNTVQVLSRFTSQGPIKPVSPNSPFGTDRNLNANR----- 543
 QY 111 QMLTMSGDLKVRKRSISGVNTATPSPVAVQSRPPVQCHPAVQKTPPVVVVKKPTP 170
 DB 544 -----GD-----TSHSPTPGKVSPLSPLS--PGIKSPTTIPAEKGNPP 581
 QY 171 TPVPVQAPVAPVTEAPFATGSSGVMQFRYPVGAATNPVVRFGTATVAGSTVTSNGM 230
 DB 582 IP-----PKKFGILTPSPSAT-----TPLTKTSGAASLTATADLASSGSSNTIV 625
 QY 231 FSGRDGDLI 239
 DB 626 ANGDVVELL 634

RESULT 25

AAE18364
 ID AAE18364 standard; Protein; 357 AA.

XX AC AAE18364;

XX DT 07-MAY-2002 (first entry)

XX DE Streptococcus pyogenes strain B514 BVH-PI mature protein.

XX BVH-PI gene; streptococcal infection; pharyngitis; erysipelas; impetigo;
 KW scarlet fever; bacteraemia; necrotising fasciitis; toxic shock; vaccine;
 KW immune response; anti-inflammatory; immunisation; antibacterial.

XX OS Streptococcus pyogenes B514.

XX PN WO200204495-A2.

XX PD 17-JAN-2002.

XX PF 06-JUL-2001; 2001WO-CA01001.

XX PR 06-JUL-2000; 2000US-216465P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Martin D, Hamel J, Brodeur B;

XX DR WPI; 2002-171701/22.

XX DR N-PSDB; AAD29299.

XX New Streptococcus pyogenes antigen useful for diagnosing, preventing or
 PT treating streptococcal infection, e.g. pharyngitis, erysipelas and
 PT impetigo, scarlet fever, and invasive diseases

XX Claim 21; Fig 16; 74pp; English.

XX The invention relates to antigens, more particularly an antigen of

CC Streptococcus pyogenes (also called group A Streptococcus (GAS))
 CC bacterial pathogen. The polypeptides and polynucleotides encoding them
 CC are useful for diagnosing, preventing or treating streptococcal
 CC infection, such as pharyngitis, erysipelas, impetigo, scarlet fever,
 CC invasive diseases (bacteraemia, necrotising fasciitis, toxic shock), and
 CC for eliciting an immune response. The polypeptides may also be used as
 CC immunogens for producing antibodies for the diagnosis and treatment of
 CC Streptococcus infection, or for passive immunisation. DNAs encoding
 CC polypeptides may also be used to design DNA probes for detecting the
 CC presence of Streptococcus in biological samples suspected of containing
 CC the bacteria. The vaccine composition is useful as a prophylactic or
 CC therapeutic treatment of Streptococcal infection in an individual
 CC susceptible to or infected with streptococcal infection. The present
 CC sequence is Streptococcus pyogenes strain B514 BVH-PI mature protein.

XX SQ Sequence 357 AA;

Query Match 7.2%; Score 120; DB 23; Length 357;
 Best Local Similarity 25.8%; Pred. No. 0.056;

Matches 61; Conservative 21; Mismatches 78; Indels 76; Gaps 11;

QY 73 YQVQSGDPTVSKIAQRYGLNMR---EIGHNNLNSY--TIYT-----QMLTMSGDLK 121
 DB 24 YTVVYGDPLSTIAAMGIDVAVLDINHIANIDIFPDTLITANYNOHGQATTL----- 77
 QY 122 VRESISGCVNTATPSP--PVAVQSRPPVQCHP-----AVQKTPPVVVVKKPTP 171
 DB 78 TVQAPASSPASVSHVPSSEPLPQASATSGPVPMAKPTPLASAKPDSSVTSASELTS 137
 QY 172 -----PPVQO--PAPVAPVTEAPFATGSSGVMQFRYPVGAATNPVVRFG 215
 DB 138 NDVSTESSESQKQPEVQEAVPFPKAAETFEVEPKTDIS-----EDPTSANRRVPNESA 192
 QY 216 TATVAG-----STVTSNGMFGSGRDGLITASNG 245
 DB 193 SEEVSAAPAQAPAEKEETSAPAAQAVADTTVATSNGLSYA--PNHAYNPMNAG 246

RESULT 26

AAE18360
 ID AAE18360 standard; Protein; 382 AA.

XX AC AAE18360;

XX DT 07-MAY-2002 (first entry)

XX DE Streptococcus pyogenes strain B514 BVH-PI protein.

XX BVH-PI gene; streptococcal infection; pharyngitis; erysipelas; impetigo;
 KW scarlet fever; bacteraemia; necrotising fasciitis; toxic shock; vaccine;
 KW immune response; anti-inflammatory; immunisation; antibacterial.

XX OS Streptococcus pyogenes B514.

XX PD 17-JAN-2002.

XX PF Key Location/Qualifiers

XX FT Peptide 1..25

XX FT Protein /label= Signal_peptide

XX FT /note= "Mature_BVH_PI_protein"

XX PN WO200204495-A2.

XX PD 17-JAN-2002.

XX PF 06-JUL-2001; 2001WO-CA01001.

XX PR 06-JUL-2000; 2000US-216465P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Martin D, Hamel J, Brodeur B;

XX DR WPI; 2002-171701/22.

XX Human, secreted protein; cancer; tumour; cardiovascular disorder;
 KW blood disorder; haemophilia; autoimmune disease; diabetes; inflammation;
 KW infection; fungal; bacterial; viral; HIV; allergy; arthritis;
 KW neurodegenerative disease; asthma; contraceptive; open reading frame;
 KW ORF.
 XX Homo sapiens.
 OS
 PN MO200011015-A1.
 PD 02-MAR-2000.
 XX
 PF 24-AUG-1999; 99WO-US19351.
 XX
 PR 24-AUG-1998; 98US-0097638.
 PR 24-AUG-1998; 98US-0097659.
 PR 03-SEP-1998; 98US-0093618.
 PR 28-SEP-1998; 98US-0102092.
 PR 25-NOV-1998; 98US-0109978.
 PR 23-DEC-1998; 98US-0113645.
 PR 23-DEC-1998; 98US-0113646.
 PR 23-AUG-1999; 99US-0379246.
 XX
 PA (ALPH-) ALPHAGEN INC.
 XX
 PI Valenzuela D, Yuan O, Hoffman H, Hall J, Raplejo P;
 DR WPI; 2000-224657/19.
 XX
 PI New secreted or transmembrane proteins and polynucleotides encoding
 PT them, useful for treating neurodegenerative disorders, autoimmune
 PT diseases and cancer -
 XX
 PS Disclosure; Page 350-351; 357pp; English.
 XX
 CC The invention relates to 40 human secreted proteins (AA94981-Y95020),
 CC and CDNA sequences encoding them (AA23423-A23462). The secreted
 CC proteins of the invention include those that are thought to be only
 CC partially secreted, i.e., transmembrane proteins. The proteins of the
 CC invention may exhibit one or more activities selected from the following:
 CC cytokine activity; cell proliferation; differentiation; immune
 CC modulation; haematopoiesis regulation; tissue growth activity;
 CC activation/inhibition activity; chemotactic/chemokinetic activity; haemostatic
 CC and thrombolytic activity; anti-inflammatory activity; and tumour
 CC inhibition activity. The proteins may be administered to patients as
 CC vaccines, and the nucleotides may be used as part of a gene therapy
 CC regimen. Diseases or conditions that may be treated using the proteins or
 CC nucleotides of the invention include autoimmune diseases; genetic
 CC disorders; haemophilia; cardiovascular diseases; cancer; bacterial,
 CC fungal and viral infections, especially HIV; multiple sclerosis;
 CC rheumatoid arthritis; pulmonary inflammation; Guillain-Barre syndrome;
 CC insulin dependent diabetes mellitus; and allergic reactions such as
 CC asthma and anaemia. They may also be used for treating wounds, burns,
 CC ulcers, osteoporosis, osteoarthritis, periodontal diseases, Alzheimer's
 CC disease, Parkinson's disease, Huntington's disease and amyotrophic
 CC lateral sclerosis (ALS). Proteins with activity/inhibition activity may
 CC additionally be useful as contraceptives. Nucleic acid sequences of the
 CC invention may be used in chromosome mapping, and as a source of the
 CC diagnostic primers and probes. Sequences AA95021-Y95023, AA95026-Y95028
 CC and AA95030 represent additional open reading frames (ORFs) encoded by
 CC the CDNA clones of the invention.
 XX
 SO Sequence 284 AA;
 Query Match 7.1%; Score 118; DB 21; Length 284;
 Best Local Similarity 39.6%; Pred. No. 0.062;
 Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

OY 132 NTAHAPSPV---AVSSAPPVQGHAYOKPPPPVVK---KPPPPVQQAAPAPPV 185
 DB 115 STVPASPSLAAVAAPSKLPEDDEPPAPPPPPPPASVPOAEPWTPA---PAPAPPS 171

OY 186 T-EAPATSSSGWMQ---FRYPGATNPVVR 212
 DB 172 TPAAPKRRSSGSVDLFLALP-AASEPVIR 201

RESULT 31
 ID AAM52321 standard; Protein; 633 AA.
 AC AAM52321,
 XX 18-JAN-2002 (first entry)
 DE Las17 protein.
 KW Actin polymerisation; Ena/VASP, vasodilator-stimulated phosphoprotein;
 KW metastatic cancer; parasitic infection; cytotoxic; Las17.
 OS Saccharomyces cerevisiae.
 XX WO200171356-A2.
 XX 27-SEP-2001.
 XX 21-MAR-2001, 2001WO-FR00843.
 XX 22-MAR-2000; 2000FR-0003637.
 XX (CNRS) CENT NAT RECH SCI.
 XX (CURT-) INST CURIE.
 XX Fradelizi J, Friederich E, Golsteyn RM, Louvard D, Noireaux V;
 PI Sykes C;
 DR WPI; 2001-639148/73.
 XX
 PT Identifying modulators of actin polymerization, potentially useful for
 PT treating tumor metastasis and parasitic infection, using proteins that
 PT contain Ena/VASP binding sites -
 XX
 PS Claim 13; Pages 105-107; 109pp; French.
 XX
 CC The present invention relates to a method for identifying modulators of
 CC actin polymerization. The method involves using proteins that contain at
 CC least one binding motif for proteins of the Ena/VASP
 CC (vasodilator-stimulated phosphoprotein) family in the preparation of
 CC reagents for identification/screening of molecules that modulate
 CC formation of the actin cytoskeleton. The proteins used in the method
 CC (i.e. the proteins with binding motif(s) for Ena/VASP proteins) do not
 CC bind to the Arp2/3 protein complex. The modulators identified by the
 CC method are potentially useful for treating disorders of actin
 CC polymerisation, e.g. metastatic cancer or parasitic infection; and as
 CC cytotoxic agents. The present sequence one such protein with binding
 CC motif(s) for Ena/VASP proteins, which was used in the method of the
 CC present invention.
 XX
 SO Sequence 633 AA;
 Query Match 7.1%; Score 118; DB 22; Length 633;
 Best Local Similarity 23.0%; Pred. No. 0.17; Indels 80; Gaps 12;
 Matches 65; Conservative 34; Mismatches 104;

OY 37 PTYNSTSGSG-----SHRTSGSGAIGSVITDSQG----- 68
 DB 329 PMRTTSGSGVRLPAPPPPPRRGAPPPPPHHTVSTNLNLSAGNSILPQATGRGPPAP 388
 OY 69 -VNRVQVGGTGVKIKQRYGLMWREIGHNNLNS-----YTYTQGWLTWMSG 118
 DB 389 PPRPARPPNPTVMQANQOQNNNSNRPRGYOTNSNNSPPPPVTFTNTLPQ-WTAATG 447
 OY 119 DLKY-----KRSLSGVTNLAHTPSPAVQSSAPPVQGHAYOKPPPPVVKKPTPP 172
 DB 448 QPAVPLPONTQAPSOATNVVAPPPPPASLQSOIPQASAPAPLPILBETTSAPPPPP 507

QY 173 P-VVOQP---APVAPPTEAPFATGSSGVMQFRYPVGTNPVRRFGTATVAGSTVSN 227
 DB 508 AFLTQPOSGAGAPAPPPPPQMP-ATSTSG-----GGSFAETT 543
 QY 228 GWMFSGRDGDLIN---ASNAGTVIQADHN-MDGASIVIQHTNG 266
 DB 544 G--DAGRDLALASIRGAGIGALRKVKDSQLDKPSVLLQEARG 584

RESULT 32
 AAG67365
 ID AAG67365 standard; Protein; 633 AA.
 AC AAG67365;
 DT 13-NOV-2001 (first entry)
 DE Amino acid sequence of a yeast Las17 protein.
 KW Wiskott-Aldrich syndrome protein; WASP; actin cytoskeleton; Las17;
 KM cell motility; actin polymerisation; cancer; parasite infection;
 KM embryonic development; immune response; wound repair.
 OS Saccharomyces cerevisiae.
 PN MO200144292-A2.
 XX 21-JUN-2001.
 PD 15-DEC-2000; 2000WO-FR03569.
 PE 16-DEC-1999; 99FR-0015900.
 PR (CNRS) CENT NAT RECH SCI.
 PA (CUR1-) INST CURIE.
 XX Noireaux V, Prost J, Sykes C, Friederich E, Golsteyn RM;
 PI Louvard D;
 DR WPI; 2001-536241/59.
 DR N-PSDB; AAH77917.
 XX New fragments of WASP family proteins, useful for detecting and
 PT identifying modulators of actin cytoskeleton formation, potential
 PT anticancer and antiparasitic agents -
 PS Claim 13; Fig 7; 162bp; French.
 CC The present sequence represents a Las17 protein. Las17 is a member of
 CC the WASP (Wiskott-Aldrich syndrome protein) family of proteins.
 CC Peptide fragments of WASP-family proteins of eukaryotic cells are used
 CC to prepare reagents for detecting compounds that inhibit or stimulate
 CC formation of the actin cytoskeleton, and thus inhibit or stimulate cell
 CC motility. The peptides are used to detect and identify compounds which
 CC are potentially useful for treating diseases associated with dysfunction
 CC of actin polymerisation, particularly metastatic cancer and parasite
 CC infection; as cytotoxic agents for inhibiting/stimulating formation of
 CC the actin cytoskeleton and for detecting side-effects, on actin
 CC polymerisation, of pharmaceuticals. By modulating actin polymerisation,
 CC these compounds affect cell motility, embryonic development, the immune
 CC response and wound repair.
 XX
 SQ Sequence 633 AA;
 Query Match 7.1%; Score 118; DB 22; Length 633;
 Best Local Similarity 23.0%; Pred. No. 0.17;
 Matches 65; Conservative 34; Mismatches 104; Indels 80; Gaps 12;
 QY 37 PTYNSTSSG-----SHRTSSGGAIGAGVITDSSG-----68
 DB 329 FMRITTESSGVRLPAPPPRRGAPAPPPPHRHVTISNTLNSAGNSLLPQATGRGAPAP 388

QY 69 -VPRNRQVKQDPTVSKIAQRYGLNWRREIGHINLINS-----YTIYTGQMLTLMSG 118
 DB 369 PPRASRSRTPTVMTQMQNQNNNSRPFQYQNTSMSSPPPPVPTFTMLTPQ-MTAAATG 447
 QY 119 DLKV-----RERSISGVNTAHTPSPVAVOSSRPVQCHPAVQKPTPPVVVKKPTPT 172
 DB 448 QPAVPLPONTQAPSGQATVVPVAPPPPPASLQSQIPQSPAPAP1PPTLPSTTSAPPPPP 507
 QY 173 P-VVOQP---APVAPPTEAPFATGSSGVMQFRYPVGTNPVRRFGTATVAGSTVSN 227
 DB 508 AFLTQPOSGAGAPAPPPPPQMP-ATSTSG-----GGSFAETT 543
 QY 228 GWMFSGRDGDLIN---ASNAGTVIQADHN-MDGASIVIQHTNG 266
 DB 544 G--DAGRDLALASIRGAGIGALRKVKDSQLDKPSVLLQEARG 584

RESULT 33
 AAY56967
 ID AAY56967 standard; Protein; 1192 AA.
 AC AAY56967;
 DT 25-APR-2000 (first entry)
 DE Human MAGI polypeptide.
 KW MAGI protein; neuroendocrine-specific protein; neuropathy; human;
 KM spinal injury; neuronal degeneration; neuromuscular disorder; cancer;
 KM psychiatric disorder; developmental disorder; inflammatory disorder;
 KM stroke; cytostatic; cerebroprotective; neuroprotective.
 OS Homo sapiens.
 PN WO200005364-A1.
 XX 03-FEB-2000.
 PD 21-JUL-1999; 99WO-GB02360.
 PE 22-JUL-1998; 98GB-0016024.
 PR 19-JUL-1999; 99GB-0016898.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 PA Michajlovich D, Prinjha RK;
 PI WPI; 2000-182693/16.
 DR N-PSDB; AAZ56886.
 XX Novel polypeptides related to neuroendocrine-specific proteins and
 PT polynucleotides useful for diagnosis of various diseases and for
 PT treatment of cancer and neurological disorders -
 PS Claim 2; Page 20-21; 35pp; English.
 CC The invention relates to human MAGI protein, which is similar to
 CC neuroendocrine-specific protein. The MAGI protein can be expressed by
 CC standard recombinant methodology. The MAGI polypeptides, polynucleotides
 CC and antibodies are useful for treating diseases, including neuropathies,
 CC spinal injury, neuronal degeneration, neuromuscular disorders,
 CC psychiatric disorders and developmental disorders, cancer, stroke and
 CC inflammatory disorders. The polynucleotide is also useful for chromosome
 CC localization and for tissue expression studies. The present sequence
 CC represents the human MAGI protein.
 XX
 SQ Sequence 1192 AA;
 Query Match 7.1%; Score 118; DB 21; Length 1192;
 Best Local Similarity 39.6%; Pred. No. 0.39;
 Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;
 QY 132 NTAHTPSFV---AVOSSRPVQCHPAVQKPTPPVVVVK---KPTPTPVVQOPAPVAPV 185

DB	Accession	Protein Name	Protein Description
DB	115	STVPAFSTLSAAVSPSKLPEDDEPPAPAPPSPAPASVQAQEPVTPPA---	PAPAPAPS 171
OY	186	T-EAEPATGSSGVQ---FRYPGATNPVVR	212
DB	172	TPAAPKRGSSGSVDETLFLALF-AASEPVIR	201
RESULT 34			
AAU04591		AAU04591 standard; Protein; 1192 AA.	
AC	AAU04591;		
XX			
DT	26-SEP-2001	(first entry)	
XX			
DE		Human Nogo protein.	
XX			
KW		Human; Nogo receptor; axonal growth; immunogen; antibody; nogo protein;	
KW		cranial trauma; cerebral trauma; spinal cord injury; stroke;	
KW		demyelinating disease; multiple sclerosis; monophasia demyelination;	
KW		encephalomyelitis; multifocal leukoencephalopathy; panencephalitis;	
KW		Marchiafava-Bignami disease; pontine myelinolysis; adrenoleukodystrophy;	
KW		Pelizaeus-Wertheimer disease; spongy degeneration; Alexander's disease;	
KW		Canavan's disease; metachromatic leukodystrophy; viral infection;	
KW		Krabbe's disease.	
XX			
OS		Homo sapiens.	
XX			
FH		Key	
FT		Domain	
FT		1034..1119	Location/Qualifiers
FT		/label="luminal extracellular domain	
FT		/note="This sequence is specifically claimed"	
FT		1055..1094	
FT		/label="Pep1	
FT		/note="Receptor binding inhibitory peptide. This	
FT		sequence is specifically claimed"	
FT		1064..1088	
FT		/label="Pep2	
FT		/note="Receptor binding inhibitory peptide. This	
FT		sequence is specifically claimed"	
FT		1074..1098	
FT		/label="Pep3	
FT		/note="Receptor binding inhibitory peptide. This	
FT		sequence is specifically claimed"	
FT		1084..1108	
FT		/label="Pep4	
FT		/note="Receptor binding inhibitory peptide. This	
FT		sequence is specifically claimed"	
FT		1095..1119	
FT		/label="Pep5	
FT		/note="Receptor binding inhibitory peptide. This	
FT		sequence is specifically claimed"	
PN	WO200151520-A2.		
XX			
PD	19-JUL-2001.		
XX			
PF	12-JAN-2001; 2001WO-US01041.		
XX			
PR	12-JAN-2000; 2000US-0175707.		
PR	26-MAY-2000; 2000US-0207356.		
PR	29-SEP-2000; 2000US-0236378.		
XX			
PA	(UYVA) UNIV YALE.		
XX			
PI	Strictmatter SM;		
XX			
DR	WPI; 2001-442138/47.		
DR	N-PSDB; AAS09453.		
XX			
FT		Novel Nogo receptor protein useful for identifying modulator of Nogo	
PT		protein or Nogo receptor protein, which is useful for treating central	

PT nervous system disorders -
XX
PS Example 1; Page 101-104; 109pp; English.
XX

The sequence is the human Nogo protein, a 250kDa myelin-associated axon growth inhibitor. The invention relates to the use of the nogo receptor CC nogo protein, their nucleic acids, vectors expressing them and antibodies against them, to isolate agents which block nogo receptor mediated axonal growth. The agent is useful for treating a central nervous system disorder which is a result of cranial or cerebral trauma, spinal cord injury, stroke or a demyelinating disease selected from multiple sclerosis, monoplasmis demyelination, encephalomyelitis, multifocal leukoencephalopathy, panencephalitis, Marchiafava-Bignami disease, pontine myelolysis, adrenoleukodystrophy, Pelizaeus-Merzhauser disease, Sporgy degeneration, Alexander's disease, Canavan's disease, metachromatic leukodystrophy, viral infection and Krabbe's disease.

SQ Sequence 1192 AA;

Query Match Best Local Similarity 7.1%; Score 118; DB 22; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6

OY 132 NTAAHPSPV---AVQSRRPVQHFAVOKRTPPVVVVK---KRPFRPPVQQPAPVAPV 185
Db 115 STVPASPPLSAVAASPSKLPEDEDFPARPPPPPAPSVASQAEPVWTFPA---PALAAPSS 171
OY 186 T-EARFATGSGVQC---FRYPVQATNPVR 212
Db 172 TPAAFKRRSGSVDETLFPALP-AASEPVIR 201

RESULT 35
AAB82349
ID AAB82349 standard; Protein; 1192 AA.
AC AAB82349;
DT 23-JUL-2001 (first entry)
XX
DE Human NOGO-A protein.
KW NOGO-A; human; chromosome 2p21; neuropathy; spinal injury;
KW brain injury; stroke; neuronal degeneration; Alzheimer's disease;
KW Parkinson's disease; neuromuscular disorder; psychiatric disorder;
KW developmental disorder; neuroprotective; motropic; neuroleptic;
KW anti-Parkinsonian; cerebroprotective; neuroleptic; diagnosis;
KW therapy.
OS Homo sapiens.
PN WO200136631-AI.
PD 25-MAY-2001.
XX
PF 14-NOV-2000; 2000WO-GB04345.
XX
PR 15-NOV-1999; 99GB-0026995.
PR 24-JAN-2000; 2000GB-0001550.
XX
PA (SMIK) SMITHLINE BEECHAM PLC.
PI Michalovich D, Prinjha R;
XX
DR WPI; 2001-34382/36.
DR N-PSDB; AAF90324.
XX
PT New polypeptide designated NOGO-C is a splice variant of the human NOGO
PT gene and may be useful in the treatment of neural disorders including
XX Alzheimer's and Parkinson's diseases -
XX Disclosure; Page 26-27; 25pp; English.

Query Match	7.1%;	Score 118;	DB 20;	Length 1708;
Best Local	20.3%;	Pred. No. 0.63;		
Matches 79;	Conservative 44;	Mismatches 120;		

```

QY      1  MVTIIANSONQKIRKILGIFEVITTCIAC--ASKPYN2STSGSHRPSGGJLA  57
Dh      548  LTAIVELTASDRLECKTIVYGNKTEFFTYVNDAHLEANGPEOYVLSFDAOSGMAGSHS  607
QY      58  I-----GSOVITDBQGVPMNRYOKQGDVVS-----KIQRGLN--  91
Dh      608  LTYELTPAGIQVAINSGNDCTATATPPGGAASPAAREVAAFCSALRYNRFTORSLTGG  667
QY      92  -WR-----EIGHT-----NNLSSTIYTGOMLTLMGSDILKVEREISISGVNTA  134
Dh      668  LMLHPGGLIGIFPPSPBGHIEWPAPFCGBCSTLYTRTWSF-----SGFSSD  713
QY      135  HTSPSPVAWSSPPVQOHPHVAOKPPPV-----  162
Dh      714  FSPD-----AAAPVLAAPLGHLPHPSPSIDWILPPSKESQVDAASVAPAPAPAGDPS  769
QY      163  VVVKKPTPPPPVQOAPAVAPVPTBAPATGSSGNQOFYPGATNPPVRRRTATVAGS  222
Dh      770  IVTLTPPLPPRKPTP--PP-----SRTRLLLYTYPBGA-----KYVAGS  809
QY      223  TVTSGNMWMSGDDGLINASAGVIAQADHNMDGASIVLOHTNGFVSSYIHIKDAQVKTG  282
Dh      810  LPESDCNW-----LVNASHPG-----HRRPGG--LCH-----AFYQGFPEAFYPTE  848
QY      283  DVTVRTGORIASKQOP--SGAALPEFRISRN  311
Dh      849  FIMREGLAAYTLTRPIIHAVADIRAEON  878

```

RESULT	38
ID	AAE18363
AC	AAE18363 standard; Protein; 364 AA.
AC	AAE18363;
DT	07-MAY-2002 (first entry)
DE	Streptococcus pyogenes strain SPY57 BVH-P1 mature protein.
FE	BVH-P1 gene; streptococcal infection; pharyngitis; erysipelas; impetigo; scarlet fever; bacteraemia; necrotising fasciitis; toxic shock; vaccine immune response; anti-inflammatory; immunisation; antibacterial.
OS	Streptococcus pyogenes SPY57.
PN	WO200204495-A2.
PD	17-JAN-2002.
PP	06-JUL-2001; 2001WO-CA01001.
PR	06-JUL-2000; 2000US-216465P.
PA	(SHIR-) SHIRE BIOCHEM INC.
PI	Martin D, Hamel J, Brodeur B;
DR	WPI; 2002-171701/22.
DR	N-PSDB; AAD29298.
FT	New Streptococcus pyogenes antigen useful for diagnosing, preventing or treating streptococcal infection, e.g. pharyngitis, erysipelas and impetigo, scarlet fever, and invasive diseases
PT	
XX	

Streptococcus pyogenes (also called group A Streptococcus (GAS)) are useful for diagnosing, preventing or treating streptococcal infection, such as pharyngitis, erysipelas, impetigo, scarlet fever, invasive diseases (bacteraemia, necrotising fasciitis, toxic shock), and for eliciting an immune response. The polypeptides may also be used as immunogens for producing antibodies for the diagnosis and treatment of Streptococcus infection, or for passive immunisation. DNAs encoding polypeptides may also be used to design DNA probes for detecting the presence of Streptococcus in biological samples suspected of containing the bacteria. The vaccine composition is useful as a prophylactic or therapeutic treatment of streptococcal infection in an individual or susceptible to or infected with streptococcal infection. The present sequence is Streptococcus pyogenes strain SPV57 BVR-pI mature protein.

Query Match	7.0%	Score 117.5;	DB 23;	Length 364;
Best Local Similarity	23.5%	Pred. No. 0.094;		
Matches 56; Conservative	22;	Mismatches 87;		

Qy	73	YQVQGDVTSKIAQRYGLNMRLEIGHINNLNSYTYTYGQWLT-----	WSGDLKRYERS	126
Db	24	YTKVYGDITSTIAEMAGIDVHVLGDINHIAITDILFPDITLTNYNQHGCAVNLTV--QAP	82	
Qy	127	ISGCVTAHTPS-----PAAVQSSRPVQOHPAV--QKTPPPVYVK-----	166	
Db	83	ASPSAVSHVPSSEELPQASATSQPTVPMAPPAIPSDVPTFPASAKPDDSVTASSELT	142	
Qy	167	-----KTPPPPVQOQPAVAPVTEAFATGSSGVAQFRPVGATPVYVR	213	
Db	143	STNDVSTLSESSQKQPEVQEAFTPKAETFEVPKTDIS-----EAPISAKRPVNE	197	
Qy	214	FGTATVAG-----STVSSNGMWSFGSGRCDLILNLSNNG	245	
Db	198	SASEEVSAAAPQAPAEKETSAIPAOKAVADLTSTVAISNGLSYA--PNAIYIPMNNG	253	

ABP25889
ID ABP25889 standard; Protein; 389 AA
XX

WT	02-JUL-2002	(first entry)
DE	Streptococcus	polypeptide SEQ ID NO. 954
XX		

group A streptococci; group B streptococcus; Streptococcus agalactiae
antiflammatory; infection; vaccine; meningitis; gene therapy.

XX WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-GB04789.
 PF
 XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 PA
 PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C,
 PI Tetteijn H;

DR WPI; 2002-352536/38.
DR N-PSDB; AEN66520.

PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -

PS Claim 1; Page 3249; 4525BP; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and anti-inflammatory
CC activity. (1), nucleic acids encoding (1), AEN6604-AEN71526 and
CC antibodies that bind (1) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (1) are used to detect Streptococcus in a
CC biological sample. (1) is used to determine whether a compound binds to
CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (1) may be used to recombinantly produce (1) and may be
CC used in gene therapy. Antibodies to (1) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

CC Sequence 389 AA;

Query Match 7.0%; Score 117.5; DB 23; Length 389;

Best Local Similarity 23.5%; Pred. No. 0.1;

Matches 56; Conservative 22; Mismatches 87; Indels 73; Gaps 8;

OY 73 YVKGQDPTVSKIAQRYGLNMRIGHINNINSYTYITGQWLT-----WSGDLKVRERS 126
DB 49 YTVKKGDTLSTIAEMGIDVHVLGINHIANIDILFPDITLITANYNGHQAATNLTV-QAP 107
OY 127 ISSGVNTAHTPS-----PVAQSSRPVQOHPAV---QKTPPVVVVK----- 166
DB 108 ASSPASVSHVPSSEPLPQASATSGPTVPMAPRATSDVPTTPFAKAPDSSVTASSETLS 167
OY 167 -----KPTTPPVVQOPAPVAPVTEAPATGSSGVMQFRYPVGAATNPVVR 213
DB 168 STNDVSTELSSSQKQPEVQEAAPTTPKAETTEVEBKTDIS-----EAPTSANRPVNE 222
OY 214 FGTATVAG-----STVTSNGMWFSGRGDGLINASNAG 245
DB 223 SASSEVSSAAPQAQAPAEKEKETSAPAAQAKAVADTTSVATSNGLSYA--PNHAYNPVNAAG 278

RESULT 40

AAE18359

ID AAE18359 standard; Protein; 389 AA.

AC AAE18359;

DT 07-MAY-2002 (first entry)

DB Streptococcus pyogenes strain SPY57 BVH-P1 protein.

KW BVH-P1 gene; streptococcal infection; pharyngitis; erysipelas; impetigo;
KW scarlet fever; bacteraemia; necrotising fasciitis; toxic shock; vaccine;
KW immune response; anti-inflammatory; immunisation; antibacterial.

OS Streptococcus pyogenes SPY57.

Key Location/Qualifiers

FT Peptide 1..25

FT Protein /label= Signal_peptide

FT /note= "Mature_BVH_P1_protein"

FN WO200204495-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-CA01001.

XX 06-JUL-2000; 2000US-216465P.

XX (SHR-) SHIRE BIOCHEM INC.

XX Martin D, Hamel J, Brodeur B;

XX WPI; 2002-171701/22.

DR N-PSDB; AAD23294.

PT New Streptococcus pyogenes antigen useful for diagnosing, preventing or
PT treating streptococcal infection, e.g. pharyngitis, erysipelas and
PT impetigo, scarlet fever, and invasive diseases -

PS Claim 21; Fig 6; 74pp; English.

CC The invention relates to antigens, more particularly an antigen of
CC Streptococcus pyogenes (also called group A Streptococcus (GAS))
CC bacterial pathogen. The polypeptides and polynucleotides encoding them
CC are useful for diagnosing, preventing or treating streptococcal
CC infection, such as pharyngitis, erysipelas, impetigo, scarlet fever,
CC invasive diseases (bacteraemia, necrotising fasciitis, toxic shock), and
CC for eliciting an immune response. The polypeptides may also be used as
CC immunogens for producing antibodies for the diagnosis and treatment of
CC Streptococcus infection, or for passive immunisation. DNAs encoding
CC polypeptides may also be used to design DNA probes for detecting the
CC presence of Streptococcus in biological samples suspected of containing
CC the bacteria. The vaccine composition is useful as a prophylactic or
CC therapeutic treatment of Streptococcal infection in an individual
CC susceptible to or infected with streptococcal infection. The present
CC sequence is Streptococcus pyogenes strain SPY57 BVH-P1 protein.

CC Sequence 389 AA;

Query Match 7.0%; Score 117.5; DB 23; Length 389;

Best Local Similarity 23.5%; Pred. No. 0.1;

Matches 56; Conservative 22; Mismatches 87; Indels 73; Gaps 8;

OY 73 YVKGQDPTVSKIAQRYGLNMRIGHINNINSYTYITGQWLT-----WSGDLKVRERS 126
DB 49 YTVKKGDTLSTIAEMGIDVHVLGINHIANIDILFPDITLITANYNGHQAATNLTV-QAP 107
OY 127 ISSGVNTAHTPS-----PVAQSSRPVQOHPAV---QKTPPVVVVK----- 166
DB 108 ASSPASVSHVPSSEPLPQASATSGPTVPMAPRATSDVPTTPFAKAPDSSVTASSETLS 167
OY 167 -----KPTTPPVVQOPAPVAPVTEAPATGSSGVMQFRYPVGAATNPVVR 213
DB 168 STNDVSTELSSSQKQPEVQEAAPTTPKAETTEVEBKTDIS-----EAPTSANRPVNE 222
OY 214 FGTATVAG-----STVTSNGMWFSGRGDGLINASNAG 245
DB 223 SASSEVSSAAPQAQAPAEKEKETSAPAAQAKAVADTTSVATSNGLSYA--PNHAYNPVNAAG 278

Search completed: July 8, 2003, 11:05:36

Job time : 74 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:04:38 ; Search time 26 Seconds
(without alignments)
364.392 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 1670
Sequence: 1 MVTIAINSQKPKIKRLGL.....LFEFRISRGVYDPLTLK 332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	209	12.5	345	US-09-405-728-5	Sequence 5, Appl1
2	151.5	9.1	478	US-08-456-6708-40	Sequence 40, Appl1
3	146.5	8.8	484	US-08-127-499A-26	Sequence 26, Appl1
4	146.5	8.8	484	US-08-482-847-26	Sequence 26, Appl1
5	130.5	7.8	437	US-08-737-716-2	Sequence 2, Appl1
6	123.5	7.4	666	US-08-737-716-14	Sequence 14, Appl1
7	121.5	7.3	232	US-08-456-6708-39	Sequence 39, Appl1
8	119.5	7.2	1426	US-09-136-574A-43	Sequence 43, Appl1
9	118	7.1	216	US-08-737-716-4	Sequence 4, Appl1
10	118	7.1	1708	US-09-462-606-2	Sequence 2, Appl1
11	111	6.6	330	US-09-134-001C-4002	Sequence 4002, Appl1
12	110.5	6.6	1751	US-09-136-574A-44	Sequence 44, Appl1
13	109.5	6.6	480	US-09-189-035-5	Sequence 5, Appl1
14	109.5	6.6	480	US-09-382-086-5	Sequence 5, Appl1
15	109.5	6.6	616	US-09-136-574A-47	Sequence 47, Appl1
16	109	6.5	907	US-08-783-774-2	Sequence 2, Appl1
17	109	6.5	907	US-09-338-599A-1	Sequence 1, Appl1
18	109	6.5	907	PCT-US95-04611A-19	Sequence 19, Appl1
19	108.5	6.5	2285	US-09-308-375-2	Sequence 2, Appl1
20	108	6.5	265	US-08-619-812-9	Sequence 9, Appl1
21	108	6.5	279	US-08-619-812-4	Sequence 4, Appl1
22	107.5	6.4	671	US-08-737-716-13	Sequence 13, Appl1
23	107	6.4	490	US-09-109-841-2	Sequence 2, Appl1
24	106.5	6.4	878	US-09-556-706B-2	Sequence 15, Appl1
25	103.5	6.2	250	US-08-155-171B-15	Sequence 15, Appl1
26	103.5	6.2	250	US-08-435-998-15	Sequence 15, Appl1
27	103	6.2	214	US-08-217-327-4	Sequence 4, Appl1

28	102	6.1	447	US-09-120-927-2	Sequence 2, Appl1
29	101.5	6.1	521	US-08-276-213-3	Sequence 3, Appl1
30	101.5	6.1	1162	US-08-728-323A-2	Sequence 2, Appl1
31	101.5	6.1	1162	US-09-298-568-2	Sequence 2, Appl1
32	100.5	6.0	1130	US-09-442-100-4	Sequence 4, Appl1
33	98.5	5.9	268	US-08-431-387-4	Sequence 4, Appl1
34	98.5	5.9	750	US-09-165-239A-4	Sequence 4, Appl1
35	98.5	5.9	1702	US-08-296-791-5	Sequence 5, Appl1
36	98.5	5.9	1702	PCT-US95-10661A-5	Sequence 5, Appl1
37	98	5.9	1693	US-08-478-507-7	Sequence 7, Appl1
38	98	5.9	1693	US-09-128-275A-7	Sequence 7, Appl1
39	98	5.9	1693	US-09-553-427-7	Sequence 7, Appl1
40	97	5.8	2972	US-09-579-181-2	Sequence 2, Appl1
41	97	5.8	3118	US-09-579-181-1	Sequence 1, Appl1
42	96	5.7	442	US-08-821-355A-5	Sequence 5, Appl1
43	96	5.7	442	US-09-003-687A-5	Sequence 5, Appl1
44	96	5.7	442	US-09-136-605-5	Sequence 5, Appl1
45	96	5.7	596	US-08-821-355A-6	Sequence 6, Appl1

ALIGNMENTS

```
RESULT 1
US-09-405-728-5
Sequence 5, Application US/09405728
Patent No. 6391316
GENERAL INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Ritchey, Anthony B.
TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMMUS
TITLE OF INVENTION: TRANSFERRIN-BINDING PROTEINS
FILE REFERENCE: 9000-0049.20
CURRENT APPLICATION NUMBER: US/09/405,728
CURRENT FILING DATE: 1999-09-24
EARLIER APPLICATION NUMBER: US 09/267,749
EARLIER FILING DATE: 1999-03-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 345
ORGANISM: Haemophilus sommus
US-09-405-728-5

Query Match 12.5%; Score 209; DB 4; Length 345;
Best Local Similarity 24.8%; Pred. No. 1.5e-11;
Matches 63; Conservative 49; Mismatches 106; Indels 36; Gaps 6;

QY 73 YQKQGDVTEKIAQRYGELNREIGHNNNSYTYITGOWLTLMSC---DLKREKISS 129
DB 120 YKVRKGGTPELIIYISGMDIKELATLNNMSEPHLSIGVLTAKNNIPDSNMTPTOTINE 179

QY 130 GVTATPTSPVAVOASRPVQHPAVOKPPPVVVVKKPPPTPVVQOPAPVAPVTEAP 189
DB 180 SEYTOVTNVE--TNNAKPTNEMQKPVATPTHTMTPINK---TEPATSNITMWP----- 229

QY 190 FATGSSGWOFRYPVGAATNPFVRFGTATVAGSVTTSNGMFFSGRGGDLINASNAQTIVQ 249
DB 230 -----TNKTIQGFSSADGG-----NKGIDISGSGQAVNAAAGRVVY 268

QY 250 ADHNMGAS--IYIOHTNPFVSSYIHKDAQVKTGPTVRCGRITASKKNOPSGAALFEFR 307
DB 269 AGPALRKYGNLITIKKHNDYLSAIVAHNESILVQDOQEVKAGQOIARKGSSGTITIKLHFE 328

QY 308 ISRNQVYVDEPLTVL 321
DB 329 IRYKGGSVDPMRVL 342

RESULT 2
US-08-456-670B-40
```

Wed Jul 9 10:00:59 2003

us-10-018-706-2.fat

Page 2

Sequence 40, Application US/08456670B
Patent No. 5932415
GENERAL INFORMATION:
APPLICANT: SCHUBERT, PETER
APPLICANT: NEUMANN, SIEGFRIED
APPLICANT: PAMELZIK, MARTINA
APPLICANT: LINKEWELER, WINFRIED
APPLICANT: BURGER, CHRISTA
APPLICANT: HOFMANN, GOTTFRIED
APPLICANT: HUBERT, ANDREAS
APPLICANT: GORBEL, STEFAN
APPLICANT: KOHLER, STEFAN
TITLE OF INVENTION: PROCESSES AND AGENTS FOR DETECTING
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLER, WHITE, ZELAND & BRANNIGAN, P.C.
STREET: 2200 CLARENDON BLVD., SUITE 1400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: US
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,670B
CLASSIFICATION: 435
FILING DATE: 01-JUN-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/412,227
FILING DATE: 27-MAR-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/075,248
FILING DATE: 11-JUN-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: DE 4239567.4
FILING DATE: 25-NOV-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: DE 421911.4
FILING DATE: 11-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: HAMLET-KING, DIANA
REGISTRATION NUMBER: 33,302
REFERENCE/DOCKET NUMBER: MERCK 1694D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-243-6333
TELEFAX: 703-243-6410
TELEX: 64191
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 478 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Listeria monocytogenes
STRAIN: EGD
US-08-456-670B-40

Query Match 9.1%, Score 151.5, DB 2, Length 478;
Best Local Similarity 21.3%, Pred. No. 5,76-06;
Matches 83, Conservative 46, Mismatches 117, Indels 143, Gaps 15;

QY 25 ITTCLILGCAKSPYNTSGSGSHRTSGGGLA--IGSQVITD----- 65
DB 102 IITSIKGG--TKYVETTESNGWTKITVNDGKGFVNGKYLDDKAVSTPAAPQEVKKEET 159
QY 66 --SGGVP-----NRQYKQGDPTVSKIAQRYG 89
DB 160 TTQGAARVAETKEVKTQTTQATTPAPVATKETETPYVDONATTHAVKSGDTLWALSKYVG 219

QY 90 LNMREIGHINNINSYTYITGQMLTMSGDLKVRERSISSGVNTAHTPSPAVAVOSSRPV 149
DB 220 VSVQDITMSNNLSS--SIYVGQKLA-----KQANTATPPAEVKT----- 259
QY 150 QCHPAVQKPTPPV-----VKKPTPPVQOPAPVAPVTE-----APPATG 193
DB 260 -EAPAEKQAAVVENNTNTNTATTEKEETATQ--QCTAKAP--TEAKPAPASTNT 313
QY 194 SSGVMQFRYPVATNPVVARFGTATVAGSTVNSGMMFSGDGLINASNGVTOADHN 253
DB 314 NNAKNTNTNTNTNTPTSKNTNTNTNTNTNTNTNANQSSNNNSASAIIAEQKH 373
QY 254 MD-----GASIVIOHTNGF--VSSYIHIDQAVKQGDPTVRTG 288
DB 374 LKAVSWGNGPPTEDCSGYTKYVPAKAGISLPRTSGAQVASTRISGSQAKPGDLY--- 430
QY 289 ORIASMKQPSGALFEF--RISRGVYV 315
DB 431 -----FQYGGSGISHVGIYV 445

RESULT 3
US-08-127-499A-26
Sequence 26, Application US/08127499A
Patent No. 5510264
GENERAL INFORMATION:
APPLICANT: VAN ALSTYNE, Diane
APPLICANT: SHARMA, Lawrence Rajendra
TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,499A
FILING DATE: 28-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 51916/102/INBI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
US-08-127-499A-26

Query Match 8.8%, Score 146.5, DB 1, Length 484;
Best Local Similarity 20.6%, Pred. No. 1,76-05;
Matches 81, Conservative 46, Mismatches 121, Indels 145, Gaps 14;

QY 25 ITTCLILGCAKSPYNTSGSGSHRTSGGGLA--IGSQVITD----- 65
DB 102 IITSIKGG--TKYVETTESNGWTKITVNDGKGFVNGKYLDDKAVSTPAAPQEVKKEET 159
QY 66 --SGGVP-----NRQYKQGDPTVSKIAQRYG 89

```

Db      160  TTQQAAPAEETKEVEKQVQTATTPAPKVAETKEFVNDQNTATTHAVXSGDTIMALSKYG  219
Qy      90  LNNREIGHNINNLSSTIYTGWLTMLMSDLKVRBRSSISGWNTHNPSPAVQSSRPV  149
        220  :::::|||||:::|||||-----KQNTATTPPAEYK-----  259
Db      150  QQHPAVOKPTPPV-----VKKPTTPPVQGPAPVAPVTEAPATSSGWCQ  199
        260  -EAPAEKQAPVYKENTNTNTATTEKETATQ--OQTAPAPTEAKPAPASTWTNA  315
Qy      200  FRAPVGA-----TNPVRRFETAVASITYSNGMFFSGEDGLIASNAGVIQ  249
        316  NKNTNTNTNTNTNTNTNTNTPSKNTVNTSNNTNTNNTSNANQSSNNSSASAILAE  375
Db      250  ADHNM-----GASIVIOHTNGF--VSSYIHKDAQVKTGDT  284
Qy      376  AQKHLCAYSWGNGEPTTFDCSGYTKCYPAKAGISLPRTSGAQVASTRISSESQAKGDL  435
Db      285  VRTGCRILASKNQPSGAALFEF--PLSRNGYV  315
        436  V-----FDDYSGSGSHVGIYV  451

```

RESULT 4
US-08-482-947-26
; Sequence 26, Application US/08482847
; Patent No. 5556757
; GENERAL INFORMATION:

APPLICANT: VAN ALSTYNE, Diane
APPLICANT: SHARMA, Lawrence Rajendra
TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,847
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/127,499
FILING DATE: 28-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 51916/104/INBI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
US-08-482-847-26

[illegible]

RESULT 5
US-08-737-716-2
; Sequence 2, Application US/08737716
; Patent No. 5955258

GENERAL INFORMATION:
 APPLICANT: GIBBE BUIST
 APPLICANT: Gerard VENEMA
 APPLICANT: Jan KOK
 APPLICANT: Adrianus Marinus LEDEROER
 TITLE OF INVENTION: process for the lysis of a culture of lactic
 TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
 TITLE OF INVENTION: lysed culture.
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
 STREET: 1100 New York Avenue, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005-3918
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Microsoft Word
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/737,716
 FILING DATE: 22-APR-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/NL95/00170
 FILING DATE: 12-MAY-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 94201353.3
 FILING DATE: 12-MAY-1994
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 437 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-737-716-2

Query Match	8.8%;	Score 146.5;	DB 1;	Length 484;
Best Local Similarity	20.6%;	Pred. No. 1.7e-05;		
Matches	81;	Conservative	46;	Mismatches 121;
				Indels 145;
				Gaps 14;

Query Match 7.8%; Score 130.5; DB 2; Length 437;

Wed Jul 9 10:00:59 2003

us-10-018-706-2.ra1

Page 4

Best Local Similarity 24.1%, Pred. No. 0.00046;
Matches 69; Conservative 39; Mismatches 119; Indels 59; Gaps 12;

Qy 26 TCCILAGCAKPTNTST-----SGSGSHRTSGSGGLAIGSVITDSQGVN- 71
Db 184 TAAUGRYAIDPSIGASINRIISQYLTFRDASASAGNTNSG---STITINNNSGINS 240
Qy 72 ---RVOYKQDPTVSKIAORYGALNREIGHINNNSYTYTQOMTLMSGDLKRENSIS 128
Db 241 SSTIYVKSQDPTLWGISQRYGISAQIOSANLNKST-IYIQQLVLVGS-----ASSTNS 295
Qy 129 SGVNTAHTPSFVAVOSSRPVQOHPAVOKPTPPVVVKKPTPT--PVPVQOPAVAPPT 186
Db 296 GGSNN-----ASTPTTSTVTPA--KPTSTTVKKSQDPTLMAISVKKTSIAOLKS 345
Qy 187 EAPATGSSGVQWQ--FRYPVGNTPVVRFGTATVAGSTVTSNGMFGSGDDGLINASN 243
Db 346 WNLHSDTYIGNLIVSQAASNP-----STGSGSTATNNNSSTSSNSASLHKV 398
Qy 244 ACTVIOADHNMDOGASIVIOHTNGFVSSYTHIKDAQVTKGDTVTCQ 289
Db 399 KQDTLWGLSQKSGSP-----ASIKAWNHL-----SSDTILLIQ 432

RESULT 6

US-08-737-716-14
; Sequence 14 Application US/08737716
; Patent No 195258
GENERAL INFORMATION:
APPLICANT: Gilde BUIST
APPLICANT: Gerard VENEMA
APPLICANT: Van KOK
TITLE OF INVENTION: Adrianus Marinus LEDERBOER
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: PCT/NL95/00170
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: EP 94201353.3
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 666 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Enterococcus hirae
IMMEDIATE SOURCE:
CLONE: Fig-5a (E. hirae)
US-08-737-716-14

Query Match 7.4%; Score 123.5; DB 2; Length 666;
Best Local Similarity 31.6%; Pred. No. 0.0038;

Matches 36; Conservative 21; Mismatches 40; Indels 17; Gaps 4;

Qy 31 AGCAKPTNTSTG---SGSHRTSGSGGLAIGSVITDSQGVNRYOYKQDPTVSKIAOR 87
Db 304 AGSSTNTGNNAASGNTSANTNTSGTQATCA-----KTVKSGDSVWIKAND 352
Qy 88 YGLAWREIGHINNNSYTYTQOMTLMSGDLKVEREISSGVNTAHTPSVA 141
Db 353 HGISMNOLIMWNTKNP-VYPQOLVYSKSSSA--SGSTNTSTGNTSSNTA 403

RESULT 7

US-08-456-670B-39
; Sequence 39 Application US/08456670B
; Patent No 5932415
GENERAL INFORMATION:
APPLICANT: SCHUBERT, PETER
APPLICANT: NEWMANN, SIEGFRIED
APPLICANT: FAMELZIK, MARTINA
APPLICANT: LINKWEILER, WINFRIED
APPLICANT: BORGER, CHRISTA
APPLICANT: HOFMANN, GOTTFRIED
APPLICANT: BOBERT, ANDREAS
APPLICANT: GOEBEL, WERNER
APPLICANT: KOHLER, STEFAN
TITLE OF INVENTION: PROCESSES AND AGENTS FOR DETECTING
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLER, WHITE, ZELANO & BRANIGAN, P.C.
STREET: 2200 CLARENDON BLVD., SUITE 1400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: US
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,670B
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/412,227
FILING DATE: 27-MAR-1995
APPLICATION NUMBER: US 08/075,248
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: DE 4239567.4
FILING DATE: 25-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 4219111.4
FILING DATE: 11-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: HAMLET-KING, DIANA
REGISTRATION NUMBER: 33,302
REFERENCE/DOCKET NUMBER: MERCK 1694D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-243-6333
TELEFAX: 703-243-6410
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Listeria monocytogenes
STRAIN: EGD


```

?      REGISTRATION NUMBER: 31_215
?      REFERENCE/DOCKET NUMBER: 1997US001/CIF
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: 215-540-9200
?      TELEFAX: 215-540-5818
?      TELEX: <Unknown>
?      INFORMATION FOR SEQ ID NO: 44:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 1751 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: protein
?      SEQUENCE DESCRIPTION: SEQ ID NO: 44:
?      US-09-136-574A-44

```

Query Match	6.6%	Score 110.5	DB 4	Length 1751
Best Local Similarity	23.3%	Pred No. 0.25		
Matches 74	Conservative 34	Mismatches 99	Indels 11	Gaps 18

[illegible]

```

RESULT 13
US-09-189-035-5
Sequence 5, Application US/09189035
Patent No. 6020165
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Quegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: CYTOKINE SIGNAL REGULATORS
FILE REFERENCE: PF-0638 US
CURRENT APPLICATION NUMBER: US/09/189, 035
CURRENT FILING DATE: 1998-11-10
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PERL Program
SEQ ID NO 5
LENGTH: 480
TYPE: PR1
ORGANISM: Homo sapiens
FEATURE: --
OTHER INFORMATION: g2245671
US-09-189-035-5

```

Query Match 6.6%; Score 109.5; DB 3; Length 480;
 Best Local Similarity 22.8%; Pred. No. 0.048;
 Matches 53; Conservative 19; Mismatches 79; Indels 81; Gaps 8;
 9 SGNQKPIRLGLIPVITTTCLLAGASAKPTYNSTSGSSHRTSSGGGLATISQVITTSQG 68

[illegible]

```

RESULT 14
US-09-382-086-5
; Sequence 5, Application US/09382086
; Patent No. 620106
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guebler, Karl J.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CYTOKINE SIGNAL REGULATORS
; FILE REFERENCE: PF-0638 US
; CURRENT APPLICATION NUMBER: US/09/382,086
; CURRENT FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 09/189,035
; EARLIER FILING DATE: 1998-11-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 480
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: g2245671
US-09-382-086-5

```

Query Match	6.6%;	Score 109.5;	DB 4;	Length 480;
Best Local Similarity	22.8%;	Pred. No. 0.048;		
Matches	53;	Conservative	19;	Mismatches 79;
			Indels	81;
			Gaps	8

```

Qy      SONQXIKRKLGLFVITITCLAGACAKPTYNSTNSGGSGSRHTSGGGLAGSOYLTDGQ 68
Db      231 SLNQRRHTSG-----SSGSGSRRENSSSSGIPAVPPS-- 267

Qy      69 VPRNRYVKQGDTVSKIAORYGLNWRIGHINNLSSTYLTYGWLTLMSGDLKVRERSIS 128
Db      268 -PTTIGPAG--SAPGSGYGMTRQLSRHSTTS-----STS 301

Qy      129 GCYNTAHTSPVAVOGSRPPVOQH-----AVQKTPPV----- 163
Db      302 SG-GYRRTSTVAOESAOPHVNGGLYSQNSISLAPPPMPQLTPQIPLTGFVARQEN 360

Qy      164 VVKKRPFP-----VYQGPAPVAPPTLAPATGSSGVWQRFYVPGATNP 209
Db      361 IADSPFPPEPPDDIPEWDDSPPEPPPVYVDEEAVALQINDVPADGP 412

```

RESULT 15
 US-09-136-574A-47
 Sequence 47, Application US/09136574A
 Patent No. 6294366
 GENERAL INFORMATION:
 APPLICANT: Farrington, Graham K
 Anderson, Paige
 Gibbs, Moreland
 Bergquist, Peter
 Daniels, Roy
 Morgan, Hugh W.

Wed Jul 9 10:00:59 2003

us-10-018-706-2.ra1

Page 8

Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 15477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ. ID NO.: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

```
Query March      6.6%; Score 109.5; DB 4; length 616;
Best Local Similarity 23.7%; Pred. No. 0.068;
Matches 58; Conservative 24; Mismatches 80; Indels 83; Gaps 13
```

DQ
34 ASKPTN-----STSGGSHR-----TSSGGGLAIGCVITLDQGVNRKYVQGEPYVK 83
|||:::||:
65 ASNVETNFVKLSGYSGADYILEVFSSGAC-----QLDPGRDND 105
||||
QY 84 IARQYLN-----WREIGHNNI--NSSTIIYTGMLTLMSGDLKRERSISSGY 131
|||||:
Db 106 IQVRFKAMWSNTYNQADBSWLQSMTNGENAKVTLY-DGIVLWG-----QEKGAV 157
|||:
QY 132 NTAHTSPVAVOSSRPVQGHAPVOKPFPVVVKKPFPFPPVQOPAPVAPPVEAP-- 189
|||:
Db 158 TPTSTPTPVSSST-----PFTPAT---PFTPTSITIIPATPATP-PTPESV 199
|||:
QY 190 -----FATGSSGVNQFRYPVGATNPVRRFGTATVAGSYVTANGMFFSGRDGLIN 240
|||:
Db 200 TDIDTDIMFLFAOGNKTIYDKDKPWLTG--VNMEGFPT--GTNVEDGWASCNLKSLAL 254
|||:
QY 241 ASNAG 245
|||:
Db 255 IANRG 259

RESULT 16
US-08-783-774-2
Sequence 2, Application us/08783774
Patent No. 6054130
GENERAL INFORMATION:
Applicant: Spaete, Richard
Inventor: Backman, Winthrop

1 TITLE OF INVENTION: NON-SPLICING VARIANTS OF
2
3 TITLE OF INVENTION: GP350/220
4 NUMBER OF SEQUENCES: 19
5
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Pennie & Edmonds
8 STREET: 1155 Avenue of the Americas
9 CITY: New York
10 STATE: NY
11
12 COUNTRY: USA
13 ZIP: 10036/2711
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Diskette
17 COMPUTER: IBM Compatible
18 OPERATING SYSTEM: DOS
19 SOFTWARE: FastSeq Version 2.0
20
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/783,774
23 FILING DATE: 15-JAN-1997
24 CLASSIFICATION: 435
25
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Coruzzi, Laura A.
28 REGISTRATION NUMBER: 30,742
29 REFERENCE/DOCKET NUMBER: 7682-037
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 212-790-9090
32 TELEFAX: 212-869-9864
33
34 TELEX: 6614 PENNIE
35
36 INFORMATION FOR SEQ ID NO: 2:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 907 amino acids
39 type: amino acid
40 STRANDEDNESS:
41 TOPOLOGY: Unknown
42
43 MOLECULE TYPE: Protein
44 US-08-783-774-2

```

Query Match      6.5%; Score 109; DB 3; Length 907;
Best Local Similarity 23.2%; Pred. No. 0.13;
Matches 66; Conservative 34; Mismatches 122; Indels 62; Gaps 13;

Oy    35 SKETVNSTGSGSHRTSGGGGLAIGSYITTSQGVPNXY--QVGQGTVSKIAQRVGLNW   92
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     430 TSPPLATTTGPADPNTTGT-----LPSTHVPTNLTPASTGETVS-----   469

Oy     93 REIGHNNINNSXYYTYTGWLT---LWSGDLKARESSISGNVTATHPSPVAVOSSRP   148
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     470 --TAADVTPPTAGTSGASPYTPSPSDPDNCTESKAPDMTSSISTPVTLPPMNTSPI-   524

Oy     149 VOHPVAOKEPF----PVVVYKKFPF--TFPVVOOPAPVPALPEAFATG-----SSGV   197
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     525 ----PAVLTPPMNAISPPIPAVTLTPPMNNTSFTLGKTSPISAVALTPPMNAISSFLTKTSP   580

Oy     198 MCPRYRV-GATNPVVRFGETAIVGSITYLSNGMFPG-CRGDDLINAAGTVLOADNMND   255
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     581 SATTTPPMNAISPTLGIK-TSPISAVALTPPMNAETPIGERTSPANNAN-----HTLG   631

Oy     256 GASVIQHMGNCVFSSSVHIHKDAQVKTEGYTVTCGRISMKNRQS   299
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     632 GIS-----PTPVYSQPKNATSAYTTGHNITSSSTSMSLRFS   670


RESULT 17
US-09-328-599A-1
Sequence 1, Application US/0932859A
Patent No. 6432679
GENERAL INFORMATION:
APPLICANT: MOND, James J. and Lees, Andrew
TITLE OF INVENTION: Enhancement of B Cell Activation by Co-Ligation of Receptors for Antigen and Complement C3d
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS: U.Sing EBV gp350/220 or EBV gp350/220 Peptide Adjuvants
ADDRESSER: Finnegan, Henderson, Farabow, Garrett &
```


Matches 63; Conservative 51; Mismatches 149; Indels 47; Gaps 11;

QY 22 FGVITTCILAGCASKRTYNSTSGSGHRTSGGGL-----AIGSOYITDSQVPRVQYK 76
Db 1398 YGIVTSTSGGTPSTSGSGYSKYSYNSAASKNNDPALIAVIOGSEGF--NAKAR 1455
QY 77 QGDTVKIAQRYGLNREIGHINNLSY-----TIYGOULTMSGDLKVRKRSISSG 130
Db 1456 SGVGANGLMQMPATKSLG--VNNMADPVQNMGGTKTLAQULEFGGVEKALAINAG 1514
QY 131 VMTAHTPSVAVOSSRPPVQOHPAVQKPTPPVVVVKKPTPPPVVQOPAPVAPVTEAPF 190
Db 1515 -----PGNVIKYGGIPPEKFNQNVYKIMANYSKLSAISST-----ASY 1555
QY 191 ATGSSGVMOFRYPVGNATNPVRRFGTATVAGSTVTSNGMWFSGRDGDLINASNAGTVIOA 250
Db 1556 YTNNSA---FR-----VSSKIQOQESGLRSHKGTDPAAKAGTAIKSLQSGKVOIA 1604
QY 251 DHMD-GASIVYQHTNGFVSYTH--IKDAQYKGTPTVTRGRIASMKNOPSGAA--LFEFR 307
Db 1605 GYSKTRAGMNVVLIKODGTVAKTMMMLTPSVYAGSVAGGQTIGKVGSTGNSGNHLIQ 1664
QY 308 ISRNGVYVDP 317
Db 1665 IEONGKTIID 1674

RESULT 20

US-08-619-812-9
; Sequence 9, Application US/08619812
; Patent No. 6100066
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: THEISEN, MICHAEL
; APPLICANT: HARLAND, RICHARD J.
; APPLICANT: RIOUX, CLEMENT R.
; TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,812
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038,719
; FILING DATE: 29-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0019.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-619-812-9

Query Match 6.5%; Score 108; DB 3; Length 265;

Best Local Similarity 22.0%; Pred. No. 0.028;
Matches 39; Conservative 35; Mismatches 69; Indels 34; Gaps 5;

QY 73 YQVKGQDTVSKIAQRYGLNREIGHINNLSYTIYGOULTMSG---DLKVRERSISS 129
Db 106 YKVRKQDTMFLIAYISGMDIKELATLNNMSEPHYLSIGOVAKIANNPDSNMIPQTINE 165
QY 130 GVNTAHTPSVAVOSSRPPVQOHPAVQKPTPPVVVVKKPTPPPVVQOPAPVAPVTEAP 189
Db 166 SEVTONTVNE--TWNANKPTNEQMKPVATPTSTWPKINK---TPPATSNIAIWP----- 215
QY 190 FATGSSGVMOFRYPVGNATNPVRRFGTATVAGSTVTSNGMWFSGRDGDLINASNAGT 246
Db 216 -----TNGKITIQGFSSADG-----NKGIDISGSRGQAVNMAAAWT 251

RESULT 21

US-08-619-812-4
; Sequence 4, Application US/08619812
; Patent No. 6100066
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: THEISEN, MICHAEL
; APPLICANT: HARLAND, RICHARD J.
; APPLICANT: RIOUX, CLEMENT R.
; TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,812
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038,719
; FILING DATE: 29-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0019.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-619-812-4

Query Match 6.5%; Score 108; DB 3; Length 279;
Best Local Similarity 22.0%; Pred. No. 0.03;

Matches 39; Conservative 35; Mismatches 69; Indels 34; Gaps 5;
QY 73 YQVKGQDTVSKIAQRYGLNREIGHINNLSYTIYGOULTMSG---DLKVRERSISS 129
Db 120 YKVRKQDTMFLIAYISGMDIKELATLNNMSEPHYLSIGOVAKIANNPDSNMIPQTINE 179
QY 130 GVNTAHTPSVAVOSSRPPVQOHPAVQKPTPPVVVVKKPTPPPVVQOPAPVAPVTEAP 189
Db 180 SEVTONTVNE--TWNANKPTNEQMKPVATPTSTWPKINK---TPPATSNIAIWP----- 229

GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: JACKMAN, WINCHROP
FILE REFERENCE: 7682-050-999
CURRENT APPLICATION NUMBER: US/09/556,706B
CURRENT FILING DATE: 2000-04-24
PRIOR FILING DATE: 1997-01-15
PRIOR APPLICATION NUMBER: 08/229,291
PRIOR FILING DATE: 1994-04-18
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patent in version 3.0
SEQ ID NO 2
LENGTH: 878
TYPE: PRT
ORGANISM: Virus
FEATURE:
OTHER INFORMATION: gp350
US-09-556-706B-2

Query Match
Best Local Similarity 22.1%; Pred. No. 0.22;
Matches 63; Conservative 29; Mismatches 118; Indels 75; Gaps 12;

DB 35 SKPTNNTSGSGSHRTSGSGGLAISQVITDSQGVNRY--QVKGDTVSKIAQRYGLNW 92
DB 430 TSPHTNTGTFADPNTTGT-----LPSSTHVPNTLTPASTGTPTS----- 469
QY 93 REIGHNNLNSSYTYTGQMLT---LWSGDLKVRERSISSGVNTAHTPSFVAVOSSRP 148
DB 470 --TADVTPTPTGTTSGASPVTPSPMDNGTESKAPDMTSTSPVTPPTFNATSP-- 524
QY 149 VOQHPAVOKPTP---PVVVVKKPTP--TPVVOQPAVAPPVTEAPFATGSSGVMOFRY 202
DB 525 ---FAVTPPTFNATSPPAVTPPTFNATSPTLGKTSPTSAVTPPTFNATSP-- 573
QY 203 PVGATNPVVRPFGTAVAGSTVTSNGMWFSGRDGDLINASVAGTVIQADHNMGAS-- 259
DB 574 -LGKTSPT-----TSAVTPPTP--ATSPFLGKTSPTSAVTPPTFNATSPVGT 619
QY 260 ---VIOHTNGFVSSYHIIHDAQVKTGDTVGTGORLASKMNOPS 299
DB 620 SFOANATNHTLGTSPTPVVTSOPKNATSAVTTGQH-----NRPS 659

RESULT 25
US-08-155-171B-15
Sequence 15, Application US/08155171B
Patent No. 5543264
GENERAL INFORMATION:
APPLICANT: Anderson, Carl W.
TITLE OF INVENTION: Co-Factor Activated Recombinant
TITLE OF INVENTION: Adenovirus Proteins (As Amended)
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/155,171B
FILING DATE: 19-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/851,217
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/545,585
FILING DATE: 29-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BNL91-01A2, AUI93-22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-155-171B-15

Query Match
Best Local Similarity 24.0%; Pred. No. 0.067;
Matches 42; Conservative 27; Mismatches 39; Indels 67; Gaps 9;

DB 82 SKIAQRYGL-----NMRREIGHNNLNSSYTYTGQMLTWSGDL----- 120
DB 7 ASLAPRHGSRPFPMGNDIGTSNMGAFS-----KSLMSGKIKRSTIKATGSAKANS 61
QY 121 -----KVRERS-----ISSGVNTAHTPSFVAVOSS-----RPVVOQ-HPAVO 156
DB 62 STGQMLRKLKEQFQKQVNDGLASGISGVVDLANQAVQKINSKIDPRPVEPPVAE 121
QY 157 KETPPVVVVKKPTP-----TPP-----VVOQPAVAPPVTEAPFATGSSG 196
DB 122 TVSPBGRKPKPRDREBTLVTDIDPPSYEBAKQGLPTTRPI--APMATGVLG 174

RESULT 26
US-08-435-998-15
Sequence 15, Application US/08435998
Patent No. 5935840
GENERAL INFORMATION:
APPLICANT: Anderson, Carl W.
TITLE OF INVENTION: Co-Factor Activated Recombinant
TITLE OF INVENTION: Adenovirus Proteins (As Amended)
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,998
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,171
FILING DATE: 19-NOV-1993
APPLICATION NUMBER: US 07/851,217
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/545,585
FILING DATE: 29-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: ENL91-01A2, AUI93-22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-435-998-15

Query Match 6.2%; Score 103.5; DB 2; Length 250;
Best Local Similarity 24.0%; Pred. No. 0.067;
Matches 42; Conservative 27; Mismatches 39; Indels 67; Gaps 9;

QY 82 SKIAQRYGL-----NMREIGHINNINSSYTYITGQMLTMSGDL-----120
DB 7 ASIAPRHGRPFMGNWDIGTSMNSGAFS-----MGLSMGINKFGSTIKNYGSKAMNS 61
QY 121 -----KREBS-----ISSGVNTAHTSPVAVQSS-----RPPVQO-HPAVQ 156
DB 62 STGQMLRDKLKEQNFQCKVVDGLASIGVVDLANQAVORKINSKLDPRPVEPEPPAVE 121
QY 157 KRPPIVVVVKKPP-----RPP-----VVOQAPVAPVTEAPFANGSSG 196
DB 122 TVSPBGRGRPRPRDRETLVTQIDEPSPYEALKQGLPTTRPI--APMATVUG 174

RESULT 27

US-08-217-327-4
Sequence 4, Application US/08217327
Patent No. 5474925
GENERAL INFORMATION:
APPLICANT: John, Maliyakal E
APPLICANT: Barton, Kenneth A
TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles and Brady
STREET: P.O. Box 2113
CITY: Madison
STATE: WI
COUNTRY: USA
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,327
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/812,233
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 1122990831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-217-327-4

Query Match 6.2%; Score 103; DB 1; Length 214;
Best Local Similarity 42.9%; Pred. No. 0.06;
Matches 27; Conservative 6; Mismatches 26; Indels 4; Gaps 3;

QY 132 NTAHTSPVAVQSSRPVQOHPAVQKTPPVVVKKPTPI-PPVQOPAPVA-PPVTEAP 189
DB 32 STPATPTPPA-STPPTTQAPPTPTATPPVSTPPTSPPPVTASPPVSTPPTSSPP 89
QY 190 PAT 192
DB 90 PAT 92

RESULT 28

US-09-120-927-2
Sequence 2, Application US/09120927
Patent No. 6262018
GENERAL INFORMATION:
APPLICANT: Kim, Jihyun Francis
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
TITLE OF INVENTION: ERMINIA AMILOVORA AND ITS USE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/120,927
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,108
FILING DATE: 06-AUG-1977
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1581
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-120-927-2

Query Match 6.1%; Score 102; DB 4; Length 447;
Best Local Similarity 23.1%; Pred. No. 0.21;
Matches 77; Conservative 51; Mismatches 147; Indels 59; Gaps 16;

QY 6 AINSQOKPKRLGL-IFGVITTCILAGCASKPTVNSTGSGSH-RTSG-----SGGL-----56
DB 30 ANSALGQOPDRQTEQMAQLAEILKSLSPSGNAATGAGNDQTTGVGNAGLNGRK 89
QY 57 -AIGSVITDQGVPRVQVQGDVYSKIAQRYGLNMRREIGHINNINSSYTI-----Y 108
DB 90 GTAGTTPGDSQ---NMLSEMGNNGLDQAITPDGGGGGIGNPLKAKMLKILARMDDQ 146
QY 109 TQQMLTMSGDLKVRERSISSGVNTAHTSPVAVQSSRPVQOHPAVQKTPPVVVKKP 168
DB 147 SDQFGQPGTGNNSASGTSSSGSGSPNDLSGKAPSGNPSGNYSPVSTFSP-----S 200

QY 169 TPTTPVVOQAPVAP-----PYTEAFATSSGVNQFRYPGATNPVVRFGTA--- 217
DB 201 TPTSPSPIDPSSPTKAGSGSTPVDHDDPVGSAG-----IDAGNSVA--FTSAGAN 251
QY 218 -TVAGSTVT-SNGMFPGRDGLINASNAGTIVIQADHNM-----DGASI--VIQHTNGF 267
DB 252 QTVLHDTITVKAGGVFDGKQFTTAGSELGDSGQSENGKPLFIEDGASLKNVTWGDG- 310
QY 268 VSSYIHI-KDAQVKTGDTVRTGQRIASMKNPQSG 300
DB 311 -ADGILYGDAKIDMLHVTNVGEDAIVYKPNAG 343

RESULT 29

US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Himmel, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3993
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276, 213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYDROTHERMAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 6.1%, Score 101.5, DB 1, Length 521;
Best Local Similarity 25.2%, Pred. No. 0.3;
Matches 53; Conservative 20; Mismatches 71; Indels 66; Gaps 12;
QY 111 QMTLMS-----GDLVRSISSGVNTAHTP-----SPVAVGSSRPVVOQHPAV 155
DB 315 QM-TFWMNDPSDGTGILKDMQVTVVDGYLAPKSIIDPVGASG-PSQSPSEV 372
QY 156 Q-KPTPPVVVKPTTPPVVOQAPV-----APPVTEAPATSSGVN----- 198
DB 373 SPSPSPSPASAKRTPTPTPTASPTPLPTATPTPTASPTSPPTASGARCASQVNS 432
QY 199 -----QEFYYPVAGANPVVRRFGTATVAGSTVTSNGMFPGRDGLINASNAGTIVIQADHNM 254

DB 433 DMGNFPTVAATN-----SGSVATKTIWVS--WTEGNGQITLNSMNAV-----TO 477
QY 255 DGASI-----YIQ-----HTNGFVSSY 271
DB 478 NGQSVTARMMSYNNVYIQPGQNTTFQFOASY 507

RESULT 30

US-08-728-323A-2
Sequence 2, Application US/08728323A
Patent No. 5918676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728, 323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-728-323A-2

Query Match 6.1%, Score 101.5, DB 2, Length 1162;
Best Local Similarity 24.0%, Pred. No. 0.95;
Matches 40; Conservative 12; Mismatches 42; Indels 73; Gaps 8;
QY 127 ISSGVNTAHTPSPVAVGSSRPVVOQHPAVQKPT-----PPVYVVKKP-----TPT 171
DB 120 VSPFTDTHSPSPALPPTQSPSSQRPPLSPFGRDSSTPMRPSPQOTTPHSPHTTP 179
QY 172 P-----PVVOQAPVAPV----- 184
DB 180 PEPSSKSSPDLAPSTLRSLKRLSSPQSPSTLNPICSP-FVSPRCDFPANSVYRPW 238
QY 185 VTAAPFATGSSGVNQFRYPGATNPVVRFGTATVA-GSTVTSNGMW 230
DB 239 AHSPIVSSS-----DGDTPP--KQPTSPISISSPSSEGSW 276
RESULT 31
US-09-298-568-2
Sequence 2, Application US/09298568
Patent No. 6322792

```

: GENERAL INFORMATION:
: APPLICANT: Kieff, Elliott D.
: APPLICANT: Bailestas, Mary E.
: APPLICANT: Kaye, Kenneth M.
: TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
: TITLE OF INVENTION: VIRUS DNA TO MEDIANE EFFICIENT EPISOME PERSISTENCE
: FILE REFERENCE: 16412-10001R
: CURRENT APPLICATION NUMBER: US/09/238,568
: CURRENT FILING DATE: 1999-04-21
: EARLIER APPLICATION NUMBER: US 60/109,422
: EARLIER FILING DATE: 1998-11-19
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 1162
: TYPE: PRNT
: ORGANISM: Kaposi's sarcoma-associated herpesvirus
: US-09-298-568-2

```

Query Match	6.1%;	Score 101.5;	DB 4;	Length 1162;
Best Local Similarity	24.0%;	Pred. No. 0.95;		
Matches 40;	Conservative 12;	Mismatches 42;	Indels 73;	Gaps 8;

[illegible]

RESULT 32
 US-09-442-100-4
 Sequence 4, Application US/09442100
 Patent No. 6359193
 GENERAL INFORMATION:
 APPLICANT: Xu, Tian
 APPLICANT: Tao, Wufan
 APPLICANT: Wang, Weiyi
 APPLICANT: Zhang, Sheng
 APPLICANT: Yu, Wan
 TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
 TITLE OF INVENTION: GENES AND METHODS BASED THEREON
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennile & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/442,100
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/411,111
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Mastrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 6523-003
 TELECOMMUNICATION INFORMATION:

```

? TELEPHONE: (212) 790-9090
? TELEFAX: (212) 869-9741/8864
?
? TELEX: 66141 PENNIE
?
? INFORMATION FOR SEQ ID NO: 4:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 1130 amino acids
?
? TYPE: amino acid
?
? TOPOLOGY: unknown
?
? MOLECULE TYPE: protein
?
US-09-442-100-4

```

Query Match	6.0%;	Score 100.5;	DB 4;	Length 1130;
Best Local Similarity	24.3%;	Pred. No. 1.1;		
Matches 50;	Conservative 29;	Mismatches 78;	Indels 49;	Gaps 10;

```

QY 22 FGVITTCIAGAGASK--PYNASTSGSC-SHRTSGSGGLAIAGQV-----ITDGGVGNRY 73
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 356 FMIHQNVNVAIGVAGVQPPPPYPLTIANQSSALQGTGSAAPSSVTYNGSIPOSMMVNRN 415
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 74 QVKGQDPTVSKIAQRYGL-----NWREIGHINNLNSYTYITYGQWLTLNSGDLKRYERS 126
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 416 -----SHNHELINISVPLQGTWPPQ-----SSSAPAQSSPSGCHIEPTMQNPINPVASNS 464
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 127 ISSGVN-----TAHTPEPV--ANQSR--PPIVQGHANVQKPR--PPVY 163
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 465 FNNPLNRRASHSANSQPSATVTATTPAPIQQPVKSMKVLKPELOTLAPTHPEMIPQPI 524
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 164 VVKKPPTPTPPVYQQAPAVAPVTEAP 189
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 525 QTVQSPFPPEGTASNVTVMPVAENP 550
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||

```

RESULT 33
 US-08-431-387-4
 Sequence 4, Application US/08431387
 Patent No. 5677163
 GENERAL INFORMATION:
 APPLICANT: Mainzer, Stanley E.
 APPLICANT: Lad, Pushkaraj J.
 APPLICANT: Schmidt, Brian
 TITLE OF INVENTION: Cleaning Compositions Containing
 TITLE OF INVENTION: No. 5677163el Alkaline Proteases
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genencor International, Inc.
 STREET: 180 Kimball Way
 CITY: South San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/431,387
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/950,856A
 FILING DATE: September 24, 1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Horn, Margaret A.
 REGISTRATION NUMBER: 33,401
 REFERENCE/DOCKET NUMBER: GC224
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 742-7216
 TELEFAX: (415) 742-7217
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 268 amino acids
 TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-431-387-4

Query Match
Best Local Similarity 24.9%; Score 98.5; DB 1; Length 268;
Pred. No. 0.22;
Matches 71; Conservative 46; Mismatches 95; Indels 73; Gaps 18;

QY 2 TVTAINSONOKFKIKLIFG-----VITTCI-----LAGCA-----SKPTNSG 44
DB 2 TVVWGISFINTQOAHNRG--IFGNGARVAVLDITGIAHPDLRLAGASISSEPSYHDNG 60
QY 45 SSGHRTSGGGLAIGSQVITDSQV-----FNRYQVQGD-----TVSKIAQRYGLNMR 93
DB 61 HGTN---VAGTIA---ALNNSIGVLGVAPRADLVAVYLDNRSGSLASVAG--GLEMA 111
QY 94 EIGHNNLSSTYTYGQWLTMSGDLKVR--ERSISSGV-----NT-----AHTP 137
DB 112 INNHHIIMWSIGSTG-----STELAVNRKANNAGILLVGAAGNTGROGVNYPARYS 165
QY 138 SPVAV-----OSSRP--VQGHAVOKPTPPVYVVKKPTPPVVOQAPVAPVTEAFA 191
DB 166 GWAIVAAVDQNGQPSFSTYGEITISAPGVANSTYGNRYVSLGTSMA-----TPHV 220
QY 192 TSSGVCMORRYVGNATNPVRRFG--TATVAGS--TVTSNGMWFSGR 234
DB 221 AGVAAVKSRYPSYTNNOIKRINQATATYLSPSLXGGLVHAQR 265

RESULT 34

US-09-165-239A-4
Sequence 4; Application US/09165239A

Patent No. 6344554
GENERAL INFORMATION:
APPLICANT: JOHNSON, ALEXANDER
APPLICANT: BRAUN, BURKHARD R
TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES FROM CANDIDA
TITLE OF INVENTION: ALBICANS ENCODING POLYPEPTIDES ASSOCIATED WITH FILAMENTOUS
FILE REFERENCE: 220022000700
CURRENT APPLICATION NUMBER: US/09/165,239A
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/068,065
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 750
TYPE: PRT
ORGANISM: Candida albicans
US-09-165-239A-4

Query Match
Best Local Similarity 5.9%; Score 98.5; DB 4; Length 750;
Pred. No. 0.96;
Matches 46; Conservative 12; Mismatches 65; Indels 23; Gaps 7;

QY 128 SSGVNTAHTSPVAVOGSSRPVVOQHAPVOKPTPPVYVVKKPTPPVVOQAPVAP--PV 185
DB 603 SSPVAPGTSSSPVAPSSAPATESAPATE--SSPVAPGTEETTPATPGAES--TPVAPVAP 659
QY 186 TEAFATGSSGVNQFRPVGATNPVVARFGTATVAGSTVSNOMVSGRGGDLINASNG 245
DB 660 SSAP-AVESSPVA---FGVETTPV-----AFVASTAKTSALVSTTEGIIPTTLESV 708
QY 246 TVIQADNMDGASIVIOHTNGFPVSY 271
DB 709 PAIOPSANS-----SYTIASVSSF 727

RESULT 35
US-08-296-791-5
Sequence 5, Application US/08296791

Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Raikov, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1702 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-5

Query Match
Best Local Similarity 5.9%; Score 98.5; DB 4; Length 1702;
Pred. No. 3.1;
Matches 42; Conservative 24; Mismatches 62; Indels 55; Gaps 8;

QY 38 TYNSTSGSGH-----RTSGGGLAIGSQVITDSQGVNRYQVQGDV 81
DB 896 TVNSLSGNGSFYVLTDLNKGDKVVTKSGTGNFTL--GVADKGTGPTKMLTLFDAS 952
QY 82 SKIAQRYGLNMRIGH-----INNLSSTYTYGQWLTMSGDLKVERISSGV 131
DB 953 N--ATRNINLVSLVGVTVDIGAWKYLRYVNRGRYDLNPE-----VEKR--NQTV 998
QY 132 NTAHTSPVAVOGSSRPVVOQHAPVOKPTPPVYVVKKPTPPVVOQAPVAPVTEAFA 191
DB 999 DTNITTPNNIQADVSV-----PSNNEIARVETPVP--PAPATPSETTEIVA 1046
QY 192 TGS 194
DB 1047 ENS 1049

RESULT 36

PCT-US95-10661A-5
Sequence 5, Application PC/TUS9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM;
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treacatin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: RP-59941/RPT
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1702 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-5

Query Match 5.9%; Score 98.5; DB 5; Length 1702;
Best Local Similarity 23.0%; Pred. No. 3.1;
Matches 42; Conservative 24; Mismatches 62; Indels 55; Gaps 8;

QY 38 TYNSTGSGSH-----RTSGGGLAIGSVITPDSQVPRRYOVKQDITV 81
DB 896 TVNSLSGNSFYITLDSNKQKGVVYTSATGFTI--QVADKTEPTKNEULTFDAS 952
QY 82 SKIAQRVGLMREIGH-----INNLSYTYTGOMLTLMGSDLKVRERSISSGV 131
DB 953 N--ATNNNLVSLVGMTVDLGAMKYLKRVNNGRYDLNPE-----VEGR--NQTV 998
QY 132 NTNHTSPVAVGSSRPVQCHPAVQKTPPVVVVKKPTTPPVVQOPAPVAPVTEAPFA 191
DB 999 DTNITTPNNIQADVPSV-----PSNNEIARVETTPVP--PAPATPSETTETVA 1046
QY 192 TGS 194
DB 1047 ENS 1049

RESULT 37
US-08-478-507-7
Sequence 7, Application US/08478507
Patent No. 6120988
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarborough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-507-7

Query Match 5.9%; Score 98; DB 3; Length 1693;
Best Local Similarity 24.9%; Pred. No. 3.4;
Matches 43; Conservative 14; Mismatches 62; Indels 54; Gaps 9;

QY 96 GHI-----NNLSYTYTGOMLTLMGSDLKVRERSISSGVNTNHTPS-PVAVGSSRPVQ 150
DB 685 GHVESANPFCGSESTLYRTW-----SEVDVSSPARPDIGFMSSEPSIP 728
QY 151 QH---PAYOKPPPVVVVKKPPTPPVQOPA-----PVA PVTAEAFATGSSGVMQFRYP 203
DB 729 SRAATPTLAPLDPDPAPDPSPPSAPALAPASGATGAPATH---QTRRRRLFTYP 785
QY 204 VGATNPVRRFGTATVAGSTVTSNGMWFSGRQDGLINASNAGTVIOADHMDG 256
DB 786 -----DSGKVFAGSLFESTCTW-----LVNASN-----VDHRRPGG 815

RESULT 38
US-09-128-275A-7
Sequence 7, Application US/09128275A
Patent No. 6229005
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarborough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6229005-A/No. 6229005-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

ADDRESSER: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/128,275A
FILING DATE: 03-AUG-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Petilthory, Joanne R.
REGISTRATION NUMBER: 42,995
REFERENCE/DOCKET NUMBER: 4600-0183.24
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0960
TELEFAX: (650) 324-0880
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-128-275A-7

Query Match 5.9%; Score 98; DB 4; Length 1693;
Best Local Similarity 24.9%; Pred. No. 3.4;
Matches 43; Conservative 14; Mismatches 62; Indels 54; Gaps 9;

QY 96 GHI-----NNLNSSTYITGQWLTLMGDLKVRERSISSGVNTAHTPS-PVAVOSSRPVQ 150
DB 685 GHVESANPFCGSESTLRTW-----SEVDAVSSPARPDLGFMSEPSIP 728
QY 151 QH---PAVOKRTPPVVVVKKPTPTPVVOQA---PVAPVTEAPFATGSSGVMOFRYP 203
DB 729 SRAATPTLAAPLPAPPADPPSPAPALAEPAAGATGAPATH---QTARHRLRLFTYP 785
QY 204 VGATNPVVRFGTAVAGSTVTSNGMFMFSGRDGLIVASVAGVVIQADHMDG 256
DB 786 -----DGSKVFAGSLFESICTW-----LVNASH-----VDHRPGG 815

RESULT 39
US-09-553-427-7
Sequence 7, Application US/09553427
Patent No. 6379891
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R

APPLICANT: Yarbough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fy, Kirk E
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
NUMBER OF SEQUENCES: No 6379891-A/No. 6379891-B Hepatitis Viral Agent
CORRESPONDENCE ADDRESS:
ADDRESSER: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/553,427
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles R.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0960
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-553-427-7

Query Match 5.9%; Score 98; DB 4; Length 1693;
Best Local Similarity 24.9%; Pred. No. 3.4;
Matches 43; Conservative 14; Mismatches 62; Indels 54; Gaps 9;

QY 96 GHI-----NNLNSSTYITGQWLTLMGDLKVRERSISSGVNTAHTPS-PVAVOSSRPVQ 150
DB 685 GHVESANPFCGSESTLRTW-----SEVDAVSSPARPDLGFMSEPSIP 728
QY 151 QH---PAVOKRTPPVVVVKKPTPTPVVOQA---PVAPVTEAPFATGSSGVMOFRYP 203
DB 729 SRAATPTLAAPLPAPPADPPSPAPALAEPAAGATGAPATH---QTARHRLRLFTYP 785

QY 204 VGATNPVRRFGTATVAGSTVTSNGMWFSGRDGLINASNAGTVIQADHNMDS 256
Db 786 -----DGSKVFAGSLFESTCTV-----LVNASN-----VDHRRPG 815

RESULT 40

US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chiviva, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match 5.8%; Score 97; DB 4; Length 2972;

Best Local Similarity 25.7%; Pred. No. 9.6; Matches 52; Conservative 20; Mismatches 52; Indels 78; Gaps 12;

QY 137 PSPVAVQS--SRPVOQHPAVQKPTPPVVVVKKPTTPPVVQOPAPVAPVTEAPFATGS 194
Db 890 PSPGLVLSGTSRP-----FTPTLSL--KTPPAVRLSPAP--PP-----GS 927
QY 195 SGVMQ-FRIVVGATNPVRRFGTATVAGSTVTSNGMWFSGRDGLINASNAGT-----V 247
Db 928 SSLKPLTVPPGYPFPAAATTTTATATTT-----AVPAPTPAPQRLI 973
QY 248 IQADHNMDSASIVLQHTNGFVSSYIHKDAQVKGTDTVTRGORTASMKNOP---SGAL 303
Db 974 LSPD-----MQLRPSGEVVSIGQ-LASLAQRVAVNAGGSKP 1009
QY 304 FEPRIERN-----GVYDPLTV 320
Db 1010 LTFQIQGNKLTLTGAQVRLAV 1031

Search completed: July 8, 2003, 11:08:55
Job time : 30 secs

THIS PAGE BLANK (USPTO)

Db 269 AGDALRYGNYLIIKHNDYSIASVAHNESILVDOOEKAGQOIAKMGSSGTWIKLHFE 328
QY 308 ISRNGVYVDPPLTVL 321
Db 329 IRYGQSVDPMRYL 342

RESULT 2

US-10-156-761-9179
Sequence 9179, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: HORIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT FILING DATE: 2002-05-29
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9179
LENGTH: 301
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9179

Query Match

Best Local Similarity 10.9%; Score 182.5; DB 9; Length 301;
Matches 83; Conservative 35; Mismatches 107; Indels 117; Gaps 15;

QY 46 GSHR-----TSGGGGLAI-----GGOVITDSQVRR----- 72
Db 5 GSHRRSPRFRSTIAVAGTGAALPLMGATGAATPTIAVSGKSAALVAAGKGA 64
QY 73 -----YOVKODPTSKAORIGLWREIGHINNINSYTTTGTGMLTMSGDKVRR 125
Db 65 EKSGETTAVRADSLSKIADEQSV-----TGGMKKLYSDN-----R 101
QY 126 SISSGVNT-----AHTPSFVAVSSRPPVOQHFAVOKPTPPVVVVKPTPTPPV 174
Db 102 SAIGGPTLIHPGLKLTIGAKSASAAATQS-----TATKPAATGVKSAAT----- 145
QY 175 VOQPAFVAPVTEAPATSSGVMOFRYPVATNPVRRFGTAVAGSTVTSNGMWFSG- 233
Db 146 APTFASKTTLTATRAADVTYTAG---YTLFVDDG-----ATIGTAVKTIAGSWSSG 192
QY 234 -----RDGDLINASNAGTVIOAD-HNMDGASIVTQHTNGFVSYYHIKDAQVKTGDT 284
Db 193 HTGVDFVVPVGTTLTIRAVAGTVASAGMGAGYENEVVRHADQYQYAHMSQLSVSTQS 252
QY 285 VRTGOR--IASMKNOPSGAAL--FEPRISHN--GVYVDPPLTVL 322
Db 253 VAEGRQLGSGATGVNVTGPHLHFEIRTPSYGSDVDPAVALR 294

RESULT 3

US-10-156-761-12234
Sequence 12234, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: HORIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT FILING DATE: 2002-05-29
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 12234
LENGTH: 203
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-12234

Query Match

Best Local Similarity 8.5%; Score 142; DB 9; Length 203;
Matches 43; Conservative 15; Mismatches 39; Indels 18; Gaps 6;

QY 226 SNGMW-----FSGRD-----GDLINASNAGTVIOADHN-----MDGASIVTQHTNGFVS 271
Db 83 AGGMAHAGSGQDPAVPSGTEVLAHGTAVKAGNAGDPAYGNATVIAKHGNGTYSQ 142
QY 272 IHIDAQVKTGDTVTRGTORIASMKN--QPSGAAL--FEPRISHN--GVYVDPPLTVL 322
Db 143 AHSRIDVRIGQVAVATGCHIALSGNTGNSGPHLHFEIRTPSYGSDVDPAVALR 197

RESULT 4

US-10-156-761-14779
Sequence 14779, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: HORIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT FILING DATE: 2002-05-29
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14779
LENGTH: 397
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14779

Query Match

Best Local Similarity 7.6%; Score 127.5; DB 9; Length 397;
Matches 79; Conservative 23; Mismatches 131; Indels 65; Gaps 14;

QY 43 SSGSHRTSGGGLAIGS---QVITDSQ--GVNRYOVKQDITVSKIA-----QRY 88
Db 141 SSGATHRADSRTRYADVAPQTTPOSGATAMETVVLHGDTLSGIADSRHVGWEGELY 200
QY 89 GLNWRIGHINNINSYTTTGTGMLTMSGDKVRRSISGVNTAHTPSVAVOSRPP 148
Db 201 AANRGITGADPD-----ILPGORILRA---KATRTDASTST-HKSTSSSKSTSK 251
QY 149 VOQPAVOKPTPPVVVVKPTPTPPVVOOPAPVAPVTEAPATSSGVMOFRYPVATN 208
Db 252 ASSDR-----TERATTSLSIV---APVNAAGTGYAHAGSS--WSKGHTGVDF 295

[illegible]

```

RESULT 5
US-09-468-147-166
: Sequence 166, Application US/09468147A
: Publication No. US20030049601A1
:
: GENERAL INFORMATION:
: APPLICANT: Abbott Laboratories
: APPLICANT: Schlauder, George G.
: APPLICANT: Erker, James C.
: APPLICANT: Desai, Suresh M.
: APPLICANT: Dawson, George J.
: APPLICANT: Mushawar, I. K.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
: TITLE OF INVENTION: HEPATITIS E VIRUS
: FILE REFERENCE: 6232.US.P1
: CURRENT APPLICATION NUMBER: US/09/468,147A
: CURRENT FILING DATE: 1999-12-21
: EARLIER APPLICATION NUMBER: US 09/117,141
: EARLIER FILING DATE: 1998-10-15
: EARLIER APPLICATION NUMBER: US 60/061,199
: EARLIER FILING DATE: 1997-10-15
: NUMBER OF SEQ ID NOS: 258
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 166
: LENGTH: 1708
:
: TYPE: PRT
:
: ORGANISM: Hepatitis E Virus
:
: FEATURE:
: OTHER INFORMATION: Xaa = Unknown or Other at position 322
: OTHER INFORMATION: Xaa = Unknown or Other at position 331
: OTHER INFORMATION: Xaa = Unknown or Other at position 445
: OTHER INFORMATION: Xaa = Unknown or Other at position 448
: OTHER INFORMATION: Xaa = Unknown or Other at position 634
: OTHER INFORMATION: Xaa = Unknown or Other at position 646
: OTHER INFORMATION: Xaa = Unknown or Other at position 811
: OTHER INFORMATION: Xaa = Unknown or Other at position 1553
: OTHER INFORMATION: Xaa = Unknown or Other at position 1578
: OTHER INFORMATION: Xaa = Unknown or Other at position 1691
:
: US-09-468-147-166

```

Query Match	7.6%	Score 127	DB 9	Length 1708
-------------	------	-----------	------	-------------

QY	27	TCIIAGC----	ASKP	YNSTSGSGSHRTSGSGGLAI-----	GSQVITDSQGVNRPYQVK	76		
Db	574	TTVVDGAHLNANGPEEYVLS	FDSA	RSQMGASHSLTVELTAPGLQKIS	SSNGLDCTATPP	633		
QY	77	QGDVYS-----	KAQRYGLN--	WR-----	ELGHI-----	NN 100		
Db	634	XGGA	PSAPGEVYAC	SALYRNRRTORSLTGGIMLHEBGLIGT	PPSPGHIESANP	693		
QY	101	INSS	YTYTGOMLT--	WSGDLKVRERSIS	SGVNTAHTPS	PVAVQSSRPV-----	149	
Db	694	FCGEGT	LYTRKMTSGFS	SDSP	PPPAAPASAAAGL	PPPT-----	PVDSIIWLPSPS 747	
QY	150	QQR-----	PAVQKP--	TPPVVVKKPT	PTPPVVOQPA	PVAPVTEA	PFATSGSGVMQ 199	
Db	748	EESH	DAVASVSP	PPAGLTSP	IVLT--	PPPPPPVRKAT	SPPPETRR-----	LL 796
QY	200	FRYP	GA	INPVRRCGT	AVAGSTY	TSNQMFM	SGRGDGLIN	SNNGTVQADHNMDCASI 259
Db	797	YTP	PGA-----	KVYAGSL	ESDCCD-----	LVNANSNP-----	HRPGG-----	831

```

QY      260  VIGHNGFVSSIIHKDAQVKTGDTVRTGQRISMKNP-SGALFERISRN 311
      : | : | : | : | : | : | : | : | : | : | : | : |
Db      832  -LCH----AFYQRFPEAFYSTEFIRREGLAAYLTTPRIIHAVAPDYRVEQN 878

```

```

RESULT 6
US-09-738-626-6241
: Sequence 6241, Application US/09738626
: Publication No. US20020197605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIJO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIAI, KEIJO
: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAOKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738, 626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 6241
: LENGTH: 205
: TYPE: PRT
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-6241

```

Query Match	7.5%;	Score 125.5;	DB 9;	Length 205;
Best Local Similarity	33.8%;	Pred. No. 0.018;		
Matches 52;	Conservative 14;	Mismatches 73;	Indels 15;	Gaps 5

```

QY      179 PVAPAPVTEAPAPATSSGVMQGRYPVG--ATNPVRRFGTATVAGSTVS----- 226
      41 ATMAPASQTDTPAGISSGVADTVAEAGVATTAVAAPATVAPANGFTTSGEPRNGTFH 1000
Db
QY      227 NGMWFSGRDGDLINASNAGTVIQADHND-GASTYIOHTNGFVSSYIHKDAQVTKGDTV 285
      101 NGIDIANSTIGTYIVAMAGTIVSSGPGASGYGOWIRIQDDGDSISTIGHMEYLYVSIGERV 160
QY      286 RTGORIASMKNP-SGAALFEFRISRNQVY-NDP 317
Db
      161 AAGQETAGMGSGGFSTGSHLHFETIIPDQVTPDP 194

```

```

RESULT 7
US-10-060-036-71
: Sequence 71, Application US/10060036
: Publication No. US20030073144A1
: GENERAL INFORMATION:
: APPLICANT: Benson, Darin R.
: APPLICANT: Kalos, Michael D.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Persing, David H.
: APPLICANT: Hepler, William T.
: APPLICANT: Jiang, Yugu
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
: FILE REFERENCE: 210121.566
: CURRENT APPLICATION NUMBER: US/10/060.036
: CURRENT FILING DATE: 2002-01-30
: NUMBER OF SEQ ID NOS: 4560
: SOFTWARE: FastSeq for Windows Version 4.0

```

Wed Jul 9 10:00:59 2003

us-10-018-706-2.rapb

Page 4

SEQ ID NO 71
LENGTH: 1192
TYPE: PRT
ORGANISM: Homo sapiens
US-10-060-036-71

Query Match
Best Local Similarity 7.1%; Score 118; DB 9; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTNHTSPV---AVQSSRPVQOHPAVOKTTPVVVK---KTPTPPVVQOPAPVAPV 185
DB 115 STVPASPLSAAVSPSKLPEDDEPPARPPPPASVSQAEPVWTPPA---PAPAAPPS 171
QY 186 T-EAPFATGSSGVWQ---FRYPVGAATNPVVR 212
DB 172 TPAAPKRRGSSGSVDETLFPALP-AASEPVIR 201

RESULT 8
US-09-789-386-2
Sequence 2, Application US/09789386
Patent No. US20020010324A1
GENERAL INFORMATION:
APPLICANT: MICHALOVICH, DAVID
APPLICANT: PRINJHA, RABINDER KUMAR
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP-30165-C1
CURRENT APPLICATION NUMBER: US/09/789,386
CURRENT FILING DATE: 2001-02-21
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: U.K. 9916898.1
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: U.K. 9816024.5
PRIOR FILING DATE: 1999-07-22
PRIOR APPLICATION NUMBER: US 09/359,208
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1192
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-789-386-2

Query Match
Best Local Similarity 7.1%; Score 118; DB 10; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTNHTSPV---AVQSSRPVQOHPAVOKTTPVVVK---KTPTPPVVQOPAPVAPV 185
DB 115 STVPASPLSAAVSPSKLPEDDEPPARPPPPASVSQAEPVWTPPA---PAPAAPPS 171
QY 186 T-EAPFATGSSGVWQ---FRYPVGAATNPVVR 212
DB 172 TPAAPKRRGSSGSVDETLFPALP-AASEPVIR 201

RESULT 9
US-09-758-140-6
Sequence 6, Application US/09758140
Patent No. US20020012965A1
GENERAL INFORMATION:
APPLICANT: Strittmatter, Stephen M.
TITLE OF INVENTION: NOGO-RECEPTOR-MEDIATED BLOCKADE OF Axonal Growth
FILE REFERENCE: 44574-5073-US
CURRENT APPLICATION NUMBER: US/09/758,140
CURRENT FILING DATE: 2001-01-12
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: US 60/175,707
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: US 60/207,366
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 60/236,378
PRIOR FILING DATE: 2000-09-29

NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-758-140-6

Query Match
Best Local Similarity 7.1%; Score 118; DB 10; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTNHTSPV---AVQSSRPVQOHPAVOKTTPVVVK---KTPTPPVVQOPAPVAPV 185
DB 115 STVPASPLSAAVSPSKLPEDDEPPARPPPPASVSQAEPVWTPPA---PAPAAPPS 171
QY 186 T-EAPFATGSSGVWQ---FRYPVGAATNPVVR 212
DB 172 TPAAPKRRGSSGSVDETLFPALP-AASEPVIR 201

RESULT 10
US-09-893-348-23
Sequence 23, Application US/09893348
Patent No. US20020072493A1
GENERAL INFORMATION:
APPLICANT: EISENBACH-SCHWARTZ, Michael
APPLICANT: COHEN, Itun R.
APPLICANT: BESERMAN, Pierre
APPLICANT: MOSONIGO, Alon
APPLICANT: MOLEMAN, Gila
TITLE OF INVENTION: ACTIVATED T-CELLS, NERVOUS SYSTEM-SPECIFIC ANTIGENS AND THEIR USES
FILE REFERENCE: EIS-SCHWARTZ-2A
CURRENT APPLICATION NUMBER: US/09/893,348
CURRENT FILING DATE: 2001-06-28
PRIOR FILING DATE: 1999-05-19
PRIOR APPLICATION NUMBER: US 09/314,161
PRIOR FILING DATE: 1999-05-19
PRIOR APPLICATION NUMBER: US 09/218,277
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: PCT/US98/14715
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: IL 124500
PRIOR FILING DATE: 1998-05-19
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.1
SEQ ID NO 23
LENGTH: 1192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-893-348-23

Query Match
Best Local Similarity 7.1%; Score 118; DB 10; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTNHTSPV---AVQSSRPVQOHPAVOKTTPVVVK---KTPTPPVVQOPAPVAPV 185
DB 115 STVPASPLSAAVSPSKLPEDDEPPARPPPPASVSQAEPVWTPPA---PAPAAPPS 171
QY 186 T-EAPFATGSSGVWQ---FRYPVGAATNPVVR 212
DB 172 TPAAPKRRGSSGSVDETLFPALP-AASEPVIR 201

RESULT 11
US-09-972-599A-6
Sequence 6, Application US/09972599A
Patent No. US2002007295A1
GENERAL INFORMATION:
APPLICANT: STRITTMATTER, STEPHEN M.
TITLE OF INVENTION: NOGO-RECEPTOR-MEDIATED BLOCKADE OF AXONAL GROWTH
FILE REFERENCE: C077 CIP US
CURRENT APPLICATION NUMBER: US/09/972,599A


```

; CURRENT FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: PCT/US01/01041
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 09/758,140
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/236,378
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/207,366
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/175,707
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1192
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-599A-6

```

```

Query Match 7.1%; Score 118; DB 10; Length 1192;
Best Local Similarity 39.6%; Pred. No. 0.62;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

```

```

QY 132 NTATPSPV---AVOSSRPVQOHAPVQKTPPVVVK---KPTPTPVVQOAPVAPV 185
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 STVPAPSPLSAAVSPKLPEDDEPPARPPPPASVSPQAPFVWTPDA---PAPAPPS 171
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 T-EAPFATGSSGVQW---FRYPVATNPVVR 212
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 TPAAPKRRGSSGVDETIFALP-AASEPVR 201
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 12
US-09-738-626-4458
; Sequence 4458, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4458
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4458

```

```

Query Match 7.0%; Score 117; DB 9; Length 237;
Best Local Similarity 22.2%; Pred. No. 0.1;
Matches 70; Conservative 44; Mismatches 93; Indels 108; Gaps 15;
QY 10 QNKPRLGLIFVITTCILAGASAKPTYNSTGSGSHRTSGGGLAISQVITDSQGV 69
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 13 QTSPTVTKGGVAFVAVAT-----GAVSTAGG-AVAAQ--ASNQPV 51
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 70 PNRVQKQDPTVAKIARVQ---LNRREIGHINNINSYTIYTGWLTLMSGDLKVRERS 126
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52 EVNFEILTANDTTLVAGSSAPQLSLAEKFPVNL-----GDQYK--- 92
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 127 ISSGVNTATPSPVANQSSRPVQOHAPVQKTPPVVVKKPTPTPVVQOAPVAPVPT 186
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 -----TIOYNADRIQADLDARGPS-----VVRPA----- 116
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 EAPFATGSSGVQWFRYPVATNPVVRFGATVAGSTVTSNGMWFSGRDGDLINASAGT 246
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 EGSYTSG-----FGA-----RMGT-----NHNQVDIANATGPIILANDGT 152
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 VIQADHNM-DGASIVIOHTNGFPVSYIHKDAQVKTGDTVTRGORTASMKNQ--PSGAL 303
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 VIDAGPASGRGMVRLQHEDGITTVYGHMETVEVTVGQTVKAGERIAGMSRGFGTSHL 212
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 -FEFRISRNGVYDP 317
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 HFEVYPAGGA-VDP 226
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
US-10-155-400-1
; Sequence 1, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANI, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-10-155-400-1

```

```

Query Match 6.7%; Score 112; DB 9; Length 957;
Best Local Similarity 24.5%; Pred. No. 1.4;
Matches 67; Conservative 38; Mismatches 104; Indels 64; Gaps 15;

```

```

QY 34 ASKPTYNSTSGSGSHR-----TSGSGGLAISG-----YTDSQGVPNRYQVQGD 79
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 677 ASGLYHSTVIGSSWSAITGVSSAVNVGFGKSPAGSSYPVAFVGTIGVTGAYRSDCG 736
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 T-----VSKIAQRVGLMWR--IGHINNINSYTIYTGWTLTMSGDL----- 120
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 737 TTWVLINDDOHQYG-NMGQAITGDHANLRVYTGNGRGIVY--GDIAGAPSSSPSSVS 793
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 KVRERSISGVNTATPSPVANQSSRPVQOHAPVQKTPPVVVKKPTP--TPPVQCP 178
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 794 PSASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 850
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 APVAPPVTE---APFATGSSGVQWFRYPVATNP-----VVRFGTATVAGSTVTS 226
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 851 SPSSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 910
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 227 NGMWFSGRDGDLINASAGTVIADHNMDCASI 259
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 911 R-YWFT-RDG-----GSSITLV--YNCWMAAI 932
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14

Wed Jul 9 10:00:59 2003

us-10-018-706-2.rapb

Page 6

```
US-09-917-384-1
; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
US-09-917-384-1
Query Match
Best Local Similarity 24.3%; Score 112; DB 9; Length 1228;
Matches 65; Conservative 20; Mismatches 92; Indels 90; Gaps 12;
QY 51 SSGGLAIGSOVITDSQGVPRYOVKQGDVTSKIAQRYGLNWEIGH-----INNIN 102
DB 992 SRGVOPEFSQATGATVAGTMTWVWQGOQTS-----WKISYVLTGATSIISND 1042
QY 103 -----SSYTIYQWLTLMSGDLKVERSISSGVNTAHTPSPVAVQSSR 146
DB 1043 LKAIFADAAARGSLNSTDLLDVEAGFEIWMQGGGLGNSFVSV-TSGTSSPTSPSPPT 1101
QY 147 PPVOQHAPVQKPTPPVVVKKPTPPVVOQAPVAPVTEAPATGSSGVMOFRYPVGA 206
DB 1102 P-----TPSP-----TPSPSPPTSPSPSPSP-----SSSGV-----ACRA 1133
QY 207 TNPVRRFGTAVGASTVTSNG-----MFGSGRDDLINASNAGTVIOADHNMDGA 257
DB 1134 TYVNSDMGSGFTATVTVTNGSRATNGMTVAMSGNQTVTNMTAL-----TOSGA 1187
QY 258 SI-----VIQ-----HTNGFVSSY 271
DB 1188 SVTATNLSYNNVIOPGOSTTFGFGNGSY 1214
RESULT 15
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
```

OTHER INFORMATION: GUXA
US-09-917-383-1

Query Match
Best Local Similarity 24.3%; Score 112; DB 9; Length 1228;
Matches 65; Conservative 20; Mismatches 92; Indels 90; Gaps 12;

```
QY 51 SSGGLAIGSOVITDSQGVPRYOVKQGDVTSKIAQRYGLNWEIGH-----INNIN 102
DB 992 SRGVOPEFSQATGATVAGTMTWVWQGOQTS-----WKISYVLTGATSIISND 1042
QY 103 -----SSYTIYQWLTLMSGDLKVERSISSGVNTAHTPSPVAVQSSR 146
DB 1043 LKAIFADAAARGSLNSTDLLDVEAGFEIWMQGGGLGNSFVSV-TSGTSSPTSPSPPT 1101
QY 147 PPVOQHAPVQKPTPPVVVKKPTPPVVOQAPVAPVTEAPATGSSGVMOFRYPVGA 206
DB 1102 P-----TPSP-----TPSPSPPTSPSPSPSP-----SSSGV-----ACRA 1133
QY 207 TNPVRRFGTAVGASTVTSNG-----MFGSGRDDLINASNAGTVIOADHNMDGA 257
DB 1134 TYVNSDMGSGFTATVTVTNGSRATNGMTVAMSGNQTVTNMTAL-----TOSGA 1187
QY 258 SI-----VIQ-----HTNGFVSSY 271
DB 1188 SVTATNLSYNNVIOPGOSTTFGFGNGSY 1214
```

```
RESULT 16
US-10-101-464A-930
; Sequence 930, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020C2
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: US/10/101,464A
; PRIOR FILING DATE: 2000-03-12
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 930
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-930
```

Query Match
Best Local Similarity 21.5%; Score 110.5; DB 9; Length 312;
Matches 52; Conservative 20; Mismatches 51; Indels 119; Gaps 9;
QY 35 SKPTVNSTSGSGSHRISGSGGLAIGSOVITDSQ-GVPRYOVKQ-----DT- 80
DB 139 ANPLTNPFD-----LQGVGVIPPIRCKSCSQAVQNGTMTLITTVVQPVDT 186
QY 81 -VSKIAQRYGLNWEIGHINNINSSYTIYQWLTLMSGDLKVERSISSGVNTAHTPSP 139
DB 187 NLSSITRKFGSDILNFKSLNGMNSTLTAYT-----T 217
QY 140 VAVQSRPPVVOQHAPVQKPTPPVVVKKPTPPVVOQAPVAPVTEAPATGSSGVMO 199
DB 218 VLVVVSQRPVLSQGVSSPPPP-----PPPPP----- 245

Query Match 6.5%; Score 108.5; DB 9; Length 474;

Best Local Similarity 20.3%; Pred. No. 1.1;

Matches 70; Conservative 49; Mismatches 123; Indels 103; Gaps 14;

QY 49 RTSGSGGLAIGSVIT-----DSQGVNRYQVQK-----GPTVSKIAQRYGLNREIGHINN 100
DB 48 KSTPRGQISTGCGVLAESSODDQGFORSYITNPTAYAPVYGLSDVYGAAGLEIGYNI 107
QY 101 LN-SSTIYTGQWLTLMGDLKXERSISGVTATPSPVAVOSSRPVQCH-----P 153
DB 108 LNSGDSLSFTSQMLDVISG-----SPTGANIETLDPNMQCTIYEQLSQSGYGVANV 160
QY 154 AVCKPPTPVV-VYKRPPTPVVQOPAPAPVTEPPTATGSGVQCFRYPGATNPVTR 212
DB 161 ALRPSGCVLMASSSYDNDQIVDPATYADAMAE---YISTEAPLNMATQESLPPGS 217
QY 213 RFGTAVAGSTVTSNGMWSGRDGLINASNACTVIQADHNDGASIVIQHTNGFVSSY- 271
DB 218 IFKITTAA-----LENGYSADSTVTA-----AAVTLPGTNTTLTNYG 257
QY 272 -----HIKDAQVKTG-----DTVR-----TGORIA-SMKQOPSG- 300
DB 258 GQTCAGGTTTLTLTAFOQLSCNTAFVETGIDVGADALRASAEDPGVQOTYSLGIDNVGGL 317
QY 301 -----AALFEFRISRGVYVDPPLTV 320
DB 318 GEIPDDALGOSIGORDVQMVNLCAAVMAGTVSNGVHMEPLYV 362

RESULT 20

US-09-932-183A-2

Sequence 2, Application US/09932183A

Patent No. US20020127641A1

GENERAL INFORMATION:

APPLICANT: Estell, David A.

TITLE OF INVENTION: Proteases From Gram-Positive Organisms

FILE REFERENCE: GC394C1-US

CURRENT APPLICATION NUMBER: US/09/932,183A

PRIOR FILING DATE: 2001-08-17

PRIOR FILING DATE: 1999-05-14

PRIOR FILING DATE: 1998-09-08

PRIOR APPLICATION NUMBER: PCT/US98/18828

PRIOR FILING DATE: 1997-09-15

NUMBER OF SEQ ID NOS: 3

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 2285

TYPE: PRT

ORGANISM: Bacillus subtilis

US-09-932-183A-2

Query Match 6.5%; Score 108.5; DB 10; Length 2285;

Best Local Similarity 20.3%; Pred. No. 7.6;

Matches 63; Conservative 51; Mismatches 149; Indels 47; Gaps 11;

QY 22 FEVITTCIACGASKPTVNSTGSGSHRTSGSG-----AIGSOVITDSQGVNRYQV 76
DB 1398 YGIIVSTSTSGGTPSTGSGSYNSAASKNVDPALIAVIOQESGF--NAKAR 1455
QY 77 QODTVSKIAQRYGLNREIGHINNLSY-----TIYTGQWLTLMGDLKXERSISG 130
DB 1456 SGVAMGLMQLMPATKSG-VNNAVDPQNMGGTKYLAQLEFGNVEALAAVYNG 1514
QY 131 VNTATPSPVAVQOHPVQOPAPVTEPPTATGSGVQCFRYPGATNPVTR 190
DB 1515 -----FQNVITRYGGLPPEKTONYKIMANYSSLSAISST-----ASY 1555
QY 191 ATGSSGVQFRYPVGAATNPVTRFETAVYAGSTVTSNGMWSGRDGLINASNACTVIOA 250
DB 1556 YNNNSA--FR-----VSSKYQOQESGLNPSHKGTDPAKAGTAKISLQSKVOIA 1604

QY 251 DHND-GASIVTQHTNGFVSSYIH-IKDAQVKTGDTVTRTGORIA-SMKQOPSGA-LFEFR 307
DB 1605 GYSTAGKMWVVIKQDDGVAKYHMLMTPSVKQSGVKAQGTIGKVGSTGNSGNHLHQ 1664

QY 308 ISRNQVYVDP 317
DB 1665 IEQNGKITDP 1674

RESULT 21

US-10-245-103-90

Sequence 90, Application US/10245103

Publication No. US20030068778A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin

APPLICANT: Eaton, Dan

APPLICANT: Filvaroff, Ellen

APPLICANT: Goddard, Audrey

APPLICANT: Grimaldi, J. Christopher

APPLICANT: Gurney, Austin

APPLICANT: Smith, Victoria

APPLICANT: Stephan, Jean-Philippe

APPLICANT: Macabe, Colin

APPLICANT: Wood, William

APPLICANT: Zhang, Zemin

APPLICANT: Fong, Sherman

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P36081C12

CURRENT APPLICATION NUMBER: US/10/245,103

PRIOR FILING DATE: 2002-09-17

PRIOR FILING DATE: 2002-07-18

PRIOR FILING DATE: 2002-07-18

PRIOR FILING DATE: 1997-09-17

PRIOR FILING DATE: 1997-09-17

PRIOR FILING DATE: 1997-11-10

PRIOR FILING DATE: 1998-03-27

PRIOR FILING DATE: 1998-05-22

PRIOR FILING DATE: 1998-06-02

PRIOR FILING DATE: 1998-06-02

PRIOR FILING DATE: 1998-06-18

PRIOR FILING DATE: 1998-06-24

PRIOR FILING DATE: 1998-06-25

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 116

SEQ ID NO 90

LENGTH: 877

TYPE: PRT

ORGANISM: Homo Sapien

US-10-245-103-90

Query Match 6.5%; Score 108; DB 9; Length 877;

Best Local Similarity 24.4%; Pred. No. 2.6;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTCIACGASKPTVNSTGSGSHRTSGSG-----GLAISQVITDSQGVNRY 73
DB 446 GADPICVAGRCLS-PGDDGLISG-RRPDCGCGGDDSTCRVLSGN-LTD-RGSPLE 500
QY 74 Q-----VKQDPTVSKIAQRYGLNREIGHINNLSY-----TIYTGQWLT----- 113
DB 501 OKLMTIPAGALRQIQOL-----RPSNYIALRGPGRSIINGNNAVDPDPG 548
QY 114 -----TMSGDLKXERSISGVT-ATPSPVAV-----QSSRPVQOHPVQKPTP 164

Db 549 RAGTFRVRRPREEKGESLSAEGPTTQPDVYMI FQENPGVYQYVSSPP---I 605
Qy 165 VKKPTTPPVVQ-QP--APVAPVTAPATGSSGVNQF-----RYPVCAIN 208
Db 606 LENTPEPPVPOLOPELRLVEPPLAPRPARPTGTLQROVRIPQMPAPHPRTPLGSPA 665
Qy 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGVW 685

RESULT 22

US-10-245-107-90
; Sequence 90, Application US/10245107
; Publication No. US20030068779A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C71
; CURRENT APPLICATION NUMBER: US/10/245,107
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-107-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

Qy 23 GVTTTILACGCAKPYNSTSGSHRTSGS-----GLATGQVITDSQGVPNRY 73
Db 446 GADPITVAGRCIS-PCDDGILISG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGSPLAGY 500
Qy 74 Q-----VKQDPTVSKIAQRYGLNWRREIGHINNINSY-----TITYGQWL----- 113
Db 501 QKILWIPAGALRIQIADL-----RPSNLYALRGPGGRSLINGNMAYVDPGGSY 548

Qy 114 ----TMSGDLKYRERSISSGVNT-AHTSPVAV-----OSSRPVQOAHVOKRTPPVVV 164
Db 549 RAGTFRVRRPREEKGESLSAEGPTTQPDVYMI FQENPGVYQYVSSPP---I 605
Qy 165 VKKPTTPPVVQ-QP--APVAPVTAPATGSSGVNQF-----RYPVCAIN 208
Db 606 LENTPEPPVPOLOPELRLVEPPLAPRPARPTGTLQROVRIPQMPAPHPRTPLGSPA 665
Qy 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGVW 685

RESULT 23

US-10-245-143-90
; Sequence 90, Application US/10245143
; Publication No. US20030068780A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C90
; CURRENT APPLICATION NUMBER: US/10/245,143
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-143-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

Qy 23 GVTTTILACGCAKPYNSTSGSHRTSGS-----GLATGQVITDSQGVPNRY 73
Db 446 GADPITVAGRCIS-PCDDGILISG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGSPLAGY 500
Qy 74 Q-----VKQDPTVSKIAQRYGLNWRREIGHINNINSY-----TITYGQWL----- 113

Wed Jul 9 10:00:59 2003

us-10-018-706-2.rapb

Page 10

Db 501 OKILMI PAGALRIQIAQL-----RPSNNYLALRGPGRSIINGNMAVDPGSGY 548
Qy 114 ----TMSGDLKXRESISGCVNT-AHTSPVAV-----OSSRPVQCHPAVOKPTPPVV 164
Db 549 RAGGTFRVYRNPREGKESLSAEGPTTQPVVYMTFOENPGVFYQVYISSPP---I 605
Qy 165 VKKPTPTPPVYO-QP--APVAPPTVEAPFATSSGVMOF-----RYPVQATN 208
Db 606 LENPTEPPVPOLOPELIRVPEPLAPAPRPARPTGTLQROVRIPOMAPPHRTPLGSPA 665
Qy 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 24
US-10-245-771-90

; Sequence 90, Application US/10245771
; Publication No. US20030068781A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Guirney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C98
; CURRENT APPLICATION NUMBER: US/10/245,771
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-771-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
Qy 23 GVITTCILAGCASKPTYNSTSGSGSHRTSGSG-----GLAIGSOVITDSQGVPNRY 73
Db 446 GAPDVCVARGCLIS-PCQDILLISG-RRPDGCVCVGGDDSTCRVLSGN--LTD-RGGPLGY 500

Qy 74 Q-----VKQGDIVSKIAORYGLNWRREIGHINNUNSSY-----TIYTGOML----- 113
Db 501 OKILMI PAGALRIQIAQL-----RPSNNYLALRGPGRSIINGNMAVDPGSGY 548
Qy 114 ----TMSGDLKXRESISGCVNT-AHTSPVAV-----OSSRPVQCHPAVOKPTPPVV 164
Db 549 RAGGTFRVYRNPREGKESLSAEGPTTQPVVYMTFOENPGVFYQVYISSPP---I 605
Qy 165 VKKPTPTPPVYO-QP--APVAPPTVEAPFATSSGVMOF-----RYPVQATN 208
Db 606 LENPTEPPVPOLOPELIRVPEPLAPAPRPARPTGTLQROVRIPOMAPPHRTPLGSPA 665
Qy 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 25
US-10-245-851-90

; Sequence 90, Application US/10245851
; Publication No. US20030068782A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Guirney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C93
; CURRENT APPLICATION NUMBER: US/10/245,851
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-851-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
Qy 23 GVITTCILAGCASKPTYNSTSGSGSHRTSGSG-----GLAIGSOVITDSQGVPNRY 73

Db 446 GARDICVAGRCLs-PGDDGILSG-RRPDGCGVGGDDSTCRVSGN--LTD-RGGLGY 500
Qy 74 Q-----VKQDPTVSKIAORYGLNREIGHINNLSY-----TIYQWL----- 113
Db 501 OKIIMIPAGLRQIOL-----RPSNYLALRPGGRSLINNGMVAVDPPGSY 548
Qy 114 -----TLMSGDLKVERSISSGVNT-AHTSPVAV-----QSSRPVQOHFAVQKPTPPVV 164
Db 549 RAGTFRVRYNRPREBKGSLSAEGPTTPVDVYMI FQENPGVYQYVSSPP---I 605
Qy 165 VKKPTPPVQV-QP--APVAPVTEAPFATGSSGVNQF-----RYPVGATN 208
Db 606 LENTPEPPVPOQLPELRPEPLAPAPRPARPTGTLQRYRIPQMPAPRPHRTPLGSPA 665
Qy 209 PVVRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGW 685

RESULT 26

US-10-245-883-90
; Sequence 90, Application US/10245883
; Publication No. US20030068783A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watande, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C70
; CURRENT APPLICATION NUMBER: US/10/245, 883
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-883-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

Qy 23 GVITTCILACAKPPTVNSTSGSHRTSGSG-----GLAISQVITDSQVPNRY 73
Db 446 GARDICVAGRCLs-PGDDGILSG-RRPDGCGVGGDDSTCRVSGN--LTD-RGGLGY 500
Qy 74 Q-----VKQDPTVSKIAORYGLNREIGHINNLSY-----TIYQWL----- 113
Db 501 OKIIMIPAGLRQIOL-----RPSNYLALRPGGRSLINNGMVAVDPPGSY 548
Qy 114 -----TLMSGDLKVERSISSGVNT-AHTSPVAV-----QSSRPVQOHFAVQKPTPPVV 164
Db 549 RAGTFRVRYNRPREBKGSLSAEGPTTPVDVYMI FQENPGVYQYVSSPP---I 605
Qy 165 VKKPTPPVQV-QP--APVAPVTEAPFATGSSGVNQF-----RYPVGATN 208
Db 606 LENTPEPPVPOQLPELRPEPLAPAPRPARPTGTLQRYRIPQMPAPRPHRTPLGSPA 665
Qy 209 PVVRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGW 685

RESULT 27

US-10-237-535-90
; Sequence 90, Application US/10237535
; Publication No. US20030073188A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watande, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C3
; CURRENT APPLICATION NUMBER: US/10/237, 535
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/091358
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/099803
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/106932
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: 60/115554
; PRIOR FILING DATE: 1999-01-12

Wed Jul 9 10:00:59 2003

us-10-018-706-2.rapb

Page 12

PRIOR APPLICATION NUMBER: 60/119342
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/123957
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: 60/123972
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: 60/127372
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: 60/131271
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/133459
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/135725
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/135729
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/135750
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/138385
PRIOR FILING DATE: 1999-06-09
PRIOR APPLICATION NUMBER: 60/140653
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/141037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/144732
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/144758
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/144790
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145228
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/146222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 60/146843
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/148188
PRIOR FILING DATE: 1999-08-10
PRIOR APPLICATION NUMBER: 60/148513
PRIOR FILING DATE: 1999-08-12
PRIOR APPLICATION NUMBER: 60/149327
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/149395
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/150114
PRIOR FILING DATE: 1999-08-20
PRIOR APPLICATION NUMBER: 60/151700
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/151734
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/177118
PRIOR FILING DATE: 2000-01-20
PRIOR APPLICATION NUMBER: 60/179851
PRIOR FILING DATE: 2000-02-02
PRIOR APPLICATION NUMBER: 60/180921
PRIOR FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: 60/187202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/198587
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: 60/199614
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 60/206330
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206368
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/209832

PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/218371
PRIOR FILING DATE: 2000-07-13
PRIOR APPLICATION NUMBER: 60/222695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: 60/229896
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/230621
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/232887
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/235147
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/261878
PRIOR FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 60/261910
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/261939
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/262150
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/26395
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 60/26421
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/267623
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/274399
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/280982
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/282129
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/282199
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/290589
PRIOR FILING DATE: 2001-05-09
PRIOR APPLICATION NUMBER: 60/180997
PRIOR FILING DATE: 1998-11-19
PRIOR APPLICATION NUMBER: 60/267213
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: 60/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 60/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 60/403297
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 60/423741
PRIOR FILING DATE: 1999-11-10
PRIOR APPLICATION NUMBER: 60/709238
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/802706
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/872035
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: 60/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/924419
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 60/927796
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 60/929404
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 60/931836
PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/941992
PRIOR FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: 60/946374
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/1001054
PRIOR FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 60/1052586
PRIOR FILING DATE: 2002-01-15

;; PRIOR APPLICATION NUMBER: 10/081056
;; PRIOR FILING DATE: 2002-02-20
;; PRIOR APPLICATION NUMBER: 10/119480
;; PRIOR FILING DATE: 2002-04-09

Query Match: 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

Qy 23 GVITTLIAGASKEPTYNSTSGSHRTSGSG-----GLAIGQVITDSQGVNRY 73
Db 446 GADPICVAGRLS-PGCDGILGSG-KRPDGGCVGCDSTCRIVSGN--LTD-RGGPLGY 500
Qy 74 Q---VKQDPTVSKIAORYGLNWRIGHINNINLSSY-----TITYGWL----- 113
Db 501 QKIMIRAGALRQIDLD-----RPSNNTLALRGPGRSITINGMAVDPPGQSY 548
Qy 114 ----TWSGDLKVERISISSGVNT-AHTPSFVAV-----QSSRPVYQHPAVQKPTPVV 164
Db 549 RAGGTVFRRYRPPREKGEKESLSAEGPTGPDVYMI FQENPGVFYQYVSSPP---I 605
Qy 165 VKKPTPPPVVQ-OP-APVAPTEAPFATGSSGVNQF-----RYPVQATN 208
Db 606 LENTPEPPVPQLOPELIRVEPPPLAPAPRAPRTGTLQROYRIPQMPAPPHPTPLGSPA 665
Qy 209 PVVRRFGTAVGSTVTSNGWM 230
Db 666 AYMKRVGHSACSAS--CGKGYM 685

RESULT 28

US-10-238-183-90
; Sequence 90. Application US/10238183
; Publication No. US20030073189A1

GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin
;; APPLICANT: Baton, Dan
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gunney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Matambe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin

;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3630R1C11

;; CURRENT APPLICATION NUMBER: US/10/238,183

;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/091358

;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/099803
;; PRIOR FILING DATE: 1998-09-10
;; PRIOR APPLICATION NUMBER: 60/106932
;; PRIOR FILING DATE: 1998-11-03
;; PRIOR APPLICATION NUMBER: 60/115554
;; PRIOR FILING DATE: 1999-01-12
;; PRIOR APPLICATION NUMBER: 60/119342
;; PRIOR FILING DATE: 1999-02-09
;; PRIOR APPLICATION NUMBER: 60/123957
;; PRIOR FILING DATE: 1999-03-12
;; PRIOR APPLICATION NUMBER: 60/123972
;; PRIOR FILING DATE: 1999-03-11
;; PRIOR APPLICATION NUMBER: 60/127372
;; PRIOR FILING DATE: 1999-04-01
;; PRIOR APPLICATION NUMBER: 60/131271
;; PRIOR FILING DATE: 1999-04-27
;; PRIOR APPLICATION NUMBER: 60/133459
;; PRIOR FILING DATE: 1999-05-11
;; PRIOR APPLICATION NUMBER: 60/135725
;; PRIOR FILING DATE: 1999-05-25
;; PRIOR APPLICATION NUMBER: 60/135729
;; PRIOR FILING DATE: 1999-05-25
;; PRIOR APPLICATION NUMBER: 60/135750
;; PRIOR FILING DATE: 1999-05-25
;; PRIOR APPLICATION NUMBER: 60/138385
;; PRIOR FILING DATE: 1999-06-09
;; PRIOR APPLICATION NUMBER: 60/140653
;; PRIOR FILING DATE: 1999-06-22
;; PRIOR APPLICATION NUMBER: 60/141037
;; PRIOR FILING DATE: 1999-06-23
;; PRIOR APPLICATION NUMBER: 60/144732
;; PRIOR FILING DATE: 1999-07-20
;; PRIOR APPLICATION NUMBER: 60/144758
;; PRIOR FILING DATE: 1999-07-20
;; PRIOR APPLICATION NUMBER: 60/144790
;; PRIOR FILING DATE: 1999-07-20
;; PRIOR APPLICATION NUMBER: 60/145228
;; PRIOR FILING DATE: 1999-07-20
;; PRIOR APPLICATION NUMBER: 60/145698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: 60/146222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: 60/146843
;; PRIOR FILING DATE: 1999-08-03
;; PRIOR APPLICATION NUMBER: 60/148188
;; PRIOR FILING DATE: 1999-08-10
;; PRIOR APPLICATION NUMBER: 60/148513
;; PRIOR FILING DATE: 1999-08-12
;; PRIOR APPLICATION NUMBER: 60/149327
;; PRIOR FILING DATE: 1999-08-17
;; PRIOR APPLICATION NUMBER: 60/149395
;; PRIOR FILING DATE: 1999-08-17
;; PRIOR APPLICATION NUMBER: 60/150114
;; PRIOR FILING DATE: 1999-08-20
;; PRIOR APPLICATION NUMBER: 60/151700
;; PRIOR FILING DATE: 1999-08-31
;; PRIOR APPLICATION NUMBER: 60/151724
;; PRIOR FILING DATE: 1999-08-31
;; PRIOR APPLICATION NUMBER: 60/162506
;; PRIOR FILING DATE: 1999-10-29
;; PRIOR APPLICATION NUMBER: 60/170262
;; PRIOR FILING DATE: 1999-12-09
;; PRIOR APPLICATION NUMBER: 60/177118
;; PRIOR FILING DATE: 2000-01-20
;; PRIOR APPLICATION NUMBER: 60/179851
;; PRIOR FILING DATE: 2000-02-02
;; PRIOR APPLICATION NUMBER: 60/180921
;; PRIOR FILING DATE: 2000-02-08
;; PRIOR APPLICATION NUMBER: 60/187202
;; PRIOR FILING DATE: 2000-03-03

PRIOR APPLICATION NUMBER: 60/198587
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: 60/199614
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 60/206330
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206368
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/209832
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/218371
PRIOR FILING DATE: 2000-07-13
PRIOR APPLICATION NUMBER: 60/222695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: 60/229896
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/230621
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/232887
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/235147
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/261878
PRIOR FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 60/261910
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/261939
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/262150
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/264395
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 60/266421
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/267623
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/274399
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/280982
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/282129
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/282199
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/290589
PRIOR FILING DATE: 2001-05-09
PRIOR APPLICATION NUMBER: 09/180997
PRIOR FILING DATE: 1998-11-19
PRIOR APPLICATION NUMBER: 09/267213
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: 09/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/403297
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 09/423741
PRIOR FILING DATE: 1999-11-10
PRIOR APPLICATION NUMBER: 09/709238
PRIOR FILING DATE: 2000-11-06
PRIOR APPLICATION NUMBER: 09/802706
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 09/872035
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 09/924419
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/927796
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 09/929404
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 09/931836

PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 09/941992
PRIOR FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: 09/946374
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 10/001054
PRIOR FILING DATE: 2001-11-10
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 10/081056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: 10/119480
PRIOR FILING DATE: 2002-04-09

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

23 GVITTCILAGCASKPTYNSTSGSHRTSGS-----GLAISQVITDQGVNRY 73
446 GAPDVCVAGRLS-PGCDGILSG-RRPDGCGGDDSTCRIVSGN-LTD-RGGLGY 500
74 Q-----VKQGDVSKIAORYGIMREIGHINNLSY-----TIYQWL----- 113
501 QKIMIPAGALRLIDIAQ-----RPSNVIATLRGPGGRSIIINGMAVDPPGSY 548
114 -----TWSGDLKVERISISGVNT-AHTSPVAV-----QSSRPVQCHPAVOKTPPVV 164
549 RAGGVFPRVNRPPRECKGSELISAEPTQPDVYMIFOEENPQVFQYVSSPP---I 605
QY 165 VKKPTPPPVVQ-QP-APVAPVTEAPFATGSGVWQF-----HYPVATN 208
DB 606 LENTPEPVVQQLPELIVRPPPLAPARPARPTGTIQRVRIQMPARPHRTPLGSPA 665
QY 209 PVRRPGATVAGSTVINSQW 230
DB 666 AYKRVGHSACSAS-CGKGVW 685

RESULT 29
US-10-238-283-90
Sequence 90, Application US/10238283
Publication No. US20030073190A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe
APPLICANT: Wacande, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C15
CURRENT APPLICATION NUMBER: US/10/238,283
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18/059114
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22

;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-238-283-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVTTCTIAGCASKPTYNSTSGSHRTSGSG-----GLAIGSVITDSQGVNRY 73
DB 446 GARDICVAGRCLS-PCGDDGILGSG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGLGY 500
QY 74 Q-----VKQGDVSKIAQRYGLNREIGHINNLSY-----TYYGQWL----- 113
DB 501 QKILMIPAGALRLQIQL-----RPSNLYALRGPGGRSTINGNMAVDPPGSY 548
QY 114 -----TMSGDLKVERISISGVNT-AHTSPVAV-----QSRPPVOQHPAVQKTPPVV 164
DB 549 RAGGTVRVYRNPREEKGESLSAEGPTQPVVYMFQENGVFYQYVSSPP--I 605
QY 165 VKKTPPTPPVQ-QP-APVAPVTEAPFATGSSGVNQF-----RYVGATN 208
DB 606 LENTPEPPVQLOPELIRVEPPLAPRPARPTGTLOQVRIQMPAPRPHPTPLGSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGMW 230
DB 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 30
US-10-238-370-90
;; Sequence 90, Application US/10238370
;; Publication No. US20030073191A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Eaton, Dan
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watanabe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3630R1C10
;; CURRENT APPLICATION NUMBER: US/10/238, 370
;; CURRENT FILING DATE: 2002-09-09
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27

;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-238-370-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVTTCTIAGCASKPTYNSTSGSHRTSGSG-----GLAIGSVITDSQGVNRY 73
DB 446 GARDICVAGRCLS-PCGDDGILGSG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGLGY 500
QY 74 Q-----VKQGDVSKIAQRYGLNREIGHINNLSY-----TYYGQWL----- 113
DB 501 QKILMIPAGALRLQIQL-----RPSNLYALRGPGGRSTINGNMAVDPPGSY 548
QY 114 -----TMSGDLKVERISISGVNT-AHTSPVAV-----QSRPPVOQHPAVQKTPPVV 164
DB 549 RAGGTVRVYRNPREEKGESLSAEGPTQPVVYMFQENGVFYQYVSSPP--I 605
QY 165 VKKTPPTPPVQ-QP-APVAPVTEAPFATGSSGVNQF-----RYVGATN 208
DB 606 LENTPEPPVQLOPELIRVEPPLAPRPARPTGTLOQVRIQMPAPRPHPTPLGSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGMW 230
DB 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 31
US-10-245-055-90
;; Sequence 90, Application US/10245055
;; Publication No. US20030073192A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Eaton, Dan
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watanabe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3630R1C88
;; CURRENT APPLICATION NUMBER: US/10/245, 055
;; CURRENT FILING DATE: 2002-09-16
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10

PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
Remaining prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 90
LENGTH: 877
TYPE: PRT
ORGANISM: Homo Sapien
US-10-245-055-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTCILAGCASKPTVNSTSGSHRTSGSG-----GLAISQVITDSQGVPNRY 73
DB 446 GADPICVAGRCLS-PCGCGILGSG-RRPDGCGVCGDDSTCRVSGN-LTD-RGGLPLG 500
QY 74 Q---VKQGDVSKIAQRYGLNWRREIGHINNLSY-----TIYGOVL----- 113
DB 501 QKIMIPAGRLQIAQL-----RPSNVLALRGCGRSIINGNNAVDPGSGY 548
QY 114 ----TLWSGDLKVEREISISGVNT-AHTSPVAV-----QSSRPVQGHAVQKPPPVV 164
DB 549 RAGGTIVRYNRPREBEKESLSAEGPTQPDVYMIPEENPGVYQVYISSPP--I 605
QY 165 VKKTPPPVVO-QP-AVAPPVTEAPPATGSSGVMP-----RYVGATN 208
DB 606 LENTPEPPVPOQLPELIVPEPLAPRPARPTGTLQOVRIPOMPAPPHRTPLGSPA 665
QY 209 PVVRRFETAVAGSTVTSNGMW 230
DB 666 AYWKRIVGHSACSAS--CGKGVW 685

RESULT 32

US-10-245-147-90
Sequence 90, Application US/10245147
Publication No. US20030073193A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Guiney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Phillippe
APPLICANT: Watande, Collin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3630R1C72
CURRENT APPLICATION NUMBER: US/10/245,147
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
Remaining prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 90
LENGTH: 877
TYPE: PRT
ORGANISM: Homo Sapien
US-10-245-147-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTCILAGCASKPTVNSTSGSHRTSGSG-----GLAISQVITDSQGVPNRY 73
DB 446 GADPICVAGRCLS-PCGCGILGSG-RRPDGCGVCGDDSTCRVSGN-LTD-RGGLPLG 500
QY 74 Q---VKQGDVSKIAQRYGLNWRREIGHINNLSY-----TIYGOVL----- 113
DB 501 QKIMIPAGRLQIAQL-----RPSNVLALRGCGRSIINGNNAVDPGSGY 548
QY 114 ----TLWSGDLKVEREISISGVNT-AHTSPVAV-----QSSRPVQGHAVQKPPPVV 164
DB 549 RAGGTIVRYNRPREBEKESLSAEGPTQPDVYMIPEENPGVYQVYISSPP--I 605
QY 165 VKKTPPPVVO-QP-AVAPPVTEAPPATGSSGVMP-----RYVGATN 208
DB 606 LENTPEPPVPOQLPELIVPEPLAPRPARPTGTLQOVRIPOMPAPPHRTPLGSPA 665
QY 209 PVVRRFETAVAGSTVTSNGMW 230
DB 666 AYWKRIVGHSACSAS--CGKGVW 685

RESULT 33

US-10-245-730-90
Sequence 90, Application US/10245730
Publication No. US20030073194A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Guiney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Phillippe
APPLICANT: Watande, Collin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3630R1C85
CURRENT APPLICATION NUMBER: US/10/245,730
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17

;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-730-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
QY 23 GVTTTCLAGCAKPTNSTSGSHRTSGSG-----GLAGSQVITDSQGVPNRY 73
DB 446 GARDICVAGRCLS-PCGDDGLISG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGLIGY 500
QY 74 Q-----VKQGPVSKIAQRYGLNREIGHINNLSY-----TITYGQWL----- 113
DB 501 QKILMTIPAGALRIQIQL-----RPSNNTALRGPGGRSTINGNMAVDPGGSY 548
QY 114 -----TLMSGDLKVERSSISGVNT-AHTPSFVAV-----QSSRPVVOCHPAVQKPTPPVV 164
DB 549 RAGTTFVRNRPREEKGESLSAEGPTTOPVDVYMIFOENPGVFQYVYISSPP---I 605
QY 165 VKKPTPPVVO-QP--APVAPVTEAPFATGSGVNOF-----RYVQATN 208
DB 606 LENTPEPPVPOQLPELIRVPPPLAPARPARPTGTLQROVRIPOMPAPPHPTPLGSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGMW 230
DB 666 AYWKRVGHSACSAS--CGKGW 685

RESULT 34

US-10-245-739-90
;; Sequence 90, Application US/10245739
;; Publication No. US20030073195A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watanabe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P363081C86
;; CURRENT FILING DATE: US/10/245, 739
;; CURRENT APPLICATION NUMBER: 2002-09-16
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18

;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-739-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
QY 23 GVTTTCLAGCAKPTNSTSGSHRTSGSG-----GLAGSQVITDSQGVPNRY 73
DB 446 GARDICVAGRCLS-PCGDDGLISG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGLIGY 500
QY 74 Q-----VKQGPVSKIAQRYGLNREIGHINNLSY-----TITYGQWL----- 113
DB 501 QKILMTIPAGALRIQIQL-----RPSNNTALRGPGGRSTINGNMAVDPGGSY 548
QY 114 -----TLMSGDLKVERSSISGVNT-AHTPSFVAV-----QSSRPVVOCHPAVQKPTPPVV 164
DB 549 RAGTTFVRNRPREEKGESLSAEGPTTOPVDVYMIFOENPGVFQYVYISSPP---I 605
QY 165 VKKPTPPVVO-QP--APVAPVTEAPFATGSGVNOF-----RYVQATN 208
DB 606 LENTPEPPVPOQLPELIRVPPPLAPARPARPTGTLQROVRIPOMPAPPHPTPLGSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGMW 230
DB 666 AYWKRVGHSACSAS--CGKGW 685

RESULT 35

US-10-246-210-90
;; Sequence 90, Application US/10246210
;; Publication No. US20030073196A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watanabe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P363081C121
;; CURRENT FILING DATE: US/10/246, 210
;; CURRENT APPLICATION NUMBER: 2002-09-18

```
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-246-210-90
```

```
Query Match      6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
```

```
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
```

```
QY 23 GVITTCILACGASKEPTYNSTSGSHRTSGSG-----GLAIGSVITDSQGVPMNY 73
DB 446 GADPICVAGRCLS-FGDDGILGSG-RRPDGCGVCGGDDSTCRIVSGN--LTD-RGGLIY 500
QY 74 Q-----VKQGPVSKIAQRYGLNREIGHINNLSY-----TIYGOVL----- 113
DB 501 QKILMIPAGLRILQIAL-----RPSNLTALRGPGRSIINGNNAVDPGSGY 548
QY 114 ----TLMSGDLKVERSISSGVNT-AHTSPYAV-----OSSRPVQCHPAVQKPPPVV 164
DB 549 RAGTIVFRYVRPREEKGESLSAEGPTTQPDVYVMI FQENPGVYGVVISPP-----I 605
QY 165 VKKPTPTPVVQ-QP--APVAPVTEAPPATGSSGVMP-----RYVGATN 208
DB 606 LENTPEPPVPOLOPEILRVEPLAPAPRARTPGTLQROVRI PQMPAPPHRTPLGSPA 665
QY 209 PIVRRFGTATVAGSTVTSNGMW 230
DB 666 AYWKRVGHSACSAS--CGKGVW 685
```

RESULT 36

US-10-239-196-90

;; Sequence 90, Application US/10239196

;; Publication No. US20030074735A1

;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin

;; APPLICANT: Eilon, Dan

;; APPLICANT: Filvaroff, Ellen

;; APPLICANT: Goddard, Audrey

;; APPLICANT: Grimaldi, J. Christopher

;; APPLICANT: Guirey, Austin

;; APPLICANT: Smith, Victoria

;; APPLICANT: Stephan, Jean-Phillippe

;; APPLICANT: Watanabe, Colin

;; APPLICANT: Wood, William

;; APPLICANT: Zhang, Zemin

;; APPLICANT: Fong, Sherman

;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

```
;; CURRENT APPLICATION NUMBER: US/10/239,196
;; CURRENT FILING DATE: 2002-09-09
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-239-196-90
```

```
Query Match      6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
```

```
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
```

```
QY 23 GVITTCILACGASKEPTYNSTSGSHRTSGSG-----GLAIGSVITDSQGVPMNY 73
DB 446 GADPICVAGRCLS-FGDDGILGSG-RRPDGCGVCGGDDSTCRIVSGN--LTD-RGGLIY 500
QY 74 Q-----VKQGPVSKIAQRYGLNREIGHINNLSY-----TIYGOVL----- 113
DB 501 QKILMIPAGLRILQIAL-----RPSNLTALRGPGRSIINGNNAVDPGSGY 548
QY 114 ----TLMSGDLKVERSISSGVNT-AHTSPYAV-----OSSRPVQCHPAVQKPPPVV 164
DB 549 RAGTIVFRYVRPREEKGESLSAEGPTTQPDVYVMI FQENPGVYGVVISPP-----I 605
QY 165 VKKPTPTPVVQ-QP--APVAPVTEAPPATGSSGVMP-----RYVGATN 208
DB 606 LENTPEPPVPOLOPEILRVEPLAPAPRARTPGTLQROVRI PQMPAPPHRTPLGSPA 665
QY 209 PIVRRFGTATVAGSTVTSNGMW 230
DB 666 AYWKRVGHSACSAS--CGKGVW 685
```

RESULT 37

US-10-243-024-90

;; Sequence 90, Application US/10243024

;; Publication No. US2003007741A1

;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin

;; APPLICANT: Eilon, Dan

;; APPLICANT: Filvaroff, Ellen

;; APPLICANT: Goddard, Audrey

;; APPLICANT: Grimaldi, J. Christopher

;; APPLICANT: Guirey, Austin

;; APPLICANT: Smith, Victoria

;; APPLICANT: Stephan, Jean-Phillippe

;; APPLICANT: Watanabe, Colin

;; APPLICANT: Wood, William

;; APPLICANT: Zhang, Zemin

;; APPLICANT: Fong, Sherman

;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

;; TITLE OF INVENTION: ACIDS ENCODING THE SAME
;; FILE REFERENCE: P3630R1C38
;; CURRENT APPLICATION NUMBER: US/10/243,024
;; CURRENT FILING DATE: 2002-09-12
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-243-024-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTCILACGAKPTYNSTSGSHRTSGS-----GLAIGSQVITDSQGVNRY 73
Db 446 GADPICVAGRLS-PCGCDGLISG-RRPDGCGVCGGDDSTCRIVSGN--LTD-RGGPLGY 500
QY 74 Q-----VKQGDVSKIAQRYGLNMRREIGHINNLSY-----TYTGQWL----- 113
Db 501 QKILMIPAGLRILQIQL-----RPSNRYALNRGPGGRSLINGMWAVDPPGSY 548
QY 114 -----TMSGDLKVERNSISSGVNT-AHTSPVAV-----QSSRPVVOQHNAVOKTTPPVV 164
Db 549 RAGGTFRYRNPREEKGESLSAEGPTTPVDVYMIFOENPGVFYQYVSSPP---I 605
QY 165 VKKPTPPPVVQ-OP--APVAPVTEAPFATGSSGVNQF-----RYVGATN 208
Db 606 LENTPEPPVPVQLOPELIRVEPLAPAPRPARPTGTLOQROVRIQMPAPRPHRTPLSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGVW 685

RESULT 38
US-10-243-409-90

;; Sequence 90, Application US/10243409
;; Publication No. US2003007742A1

;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin

;; APPLICANT: Eaton, Dan

;; APPLICANT: Filvaroff, Ellen

;; APPLICANT: Goddard, Audrey

;; APPLICANT: Grimaldi, J. Christopher

;; APPLICANT: Guiney, Austin

;; APPLICANT: Smith, Victoria

;; APPLICANT: Stephan, Jean-Phillippe

;; APPLICANT: Watanabe, Colin

;; APPLICANT: Wood, William

;; APPLICANT: Zhang, Zemin

;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3630R1C41
;; CURRENT APPLICATION NUMBER: US/10/243,409
;; CURRENT FILING DATE: 2002-09-12
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-243-409-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTCILACGAKPTYNSTSGSHRTSGS-----GLAIGSQVITDSQGVNRY 73
Db 446 GADPICVAGRLS-PCGCDGLISG-RRPDGCGVCGGDDSTCRIVSGN--LTD-RGGPLGY 500
QY 74 Q-----VKQGDVSKIAQRYGLNMRREIGHINNLSY-----TYTGQWL----- 113
Db 501 QKILMIPAGLRILQIQL-----RPSNRYALNRGPGGRSLINGMWAVDPPGSY 548
QY 114 -----TMSGDLKVERNSISSGVNT-AHTSPVAV-----QSSRPVVOQHNAVOKTTPPVV 164
Db 549 RAGGTFRYRNPREEKGESLSAEGPTTPVDVYMIFOENPGVFYQYVSSPP---I 605
QY 165 VKKPTPPPVVQ-OP--APVAPVTEAPFATGSSGVNQF-----RYVGATN 208
Db 606 LENTPEPPVPVQLOPELIRVEPLAPAPRPARPTGTLOQROVRIQMPAPRPHRTPLSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGVW 685

RESULT 39
US-10-245-033-90

;; Sequence 90, Application US/10245033
;; Publication No. US20030078401A1

;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin

;; APPLICANT: Eaton, Dan

;; APPLICANT: Filvaroff, Ellen

;; APPLICANT: Goddard, Audrey

;; APPLICANT: Grimaldi, J. Christopher

;; APPLICANT: Guiney, Austin

;; APPLICANT: Smith, Victoria

;; APPLICANT: Stephan, Jean-Phillippe

;; APPLICANT: Watanabe, Colin

```
/ APPLICANT: Wood,William
/ APPLICANT: Zhang,Zemin
/ APPLICANT: Fong,Sherman
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3630R1C84
/ CURRENT APPLICATION NUMBER: US/10/245,033
/ CURRENT FILING DATE: 2002-09-16
/ PRIOR APPLICATION NUMBER: 10/197942
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/059114
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/063046
/ PRIOR FILING DATE: 1997-10-24
/ PRIOR APPLICATION NUMBER: 60/065027
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/086478
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/090557
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090689
/ PRIOR FILING DATE: 1998-06-25
/ PRIOR APPLICATION NUMBER: 60/090689
/ NUMBER OF SEQ ID NOS: 116
/ Remaining Prior Application data removed - See file wrapper or PALM.
/ SEQ ID NO 90
/ LENGTH: 877
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-245-033-90

Query Match      6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVTTTCILAGCASKPTYNSTSGSGSHRTSGSG-----GLAIGSVITDSQGVNRY 73
DB 446 GAPDICVAGRLCL-S-PGCDGILGSG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGPLGY 500
QY 74 Q-----VKQGDVYSKIQRYGLNMRREIGHINNLSY-----TITYGWL----- 113
DB 501 OKILMIPAGALRLQIQL-----RPSNLTALRGPGGRSLINGNNAVDPPGSY 548
QY 114 -----TWSGDLKVERISISSGVNT-AHTPSPAV-----OSSRPVQCHPAVOKPTPPVV 164
DB 549 RAGGTATFRNRPREGKESLSABGPTTOPVDVYMIFOENRGVYQYVYISSPP---I 605
QY 165 VKKPTTPPVVQ-QP--APVAPVTEAPFATGSSGVNQF-----RYPGAIN 208
DB 606 LENPTPEPVPOLOPELIRVEPPLAPRPARTGTILQROVRIQMPAPRPHPTPLGSPA 665
QY 209 PVVRPFGTATVAGSTVTNSGMW 230
DB 666 AYWKRVGHSACSAS--CGKGVW 685

RESULT 40
US-10-245-621-90
/ Sequence 90, Application US/10245621
/ Publication No. US20030077743A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin
/ APPLICANT: Eaton, Dan
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin
/ APPLICANT: Smith, Victoria
```

```
/ APPLICANT: Stephan,Jean-Phillipe
/ APPLICANT: Watanabe,Colin
/ APPLICANT: Wood,William
/ APPLICANT: Zhang,Zemin
/ APPLICANT: Fong,Sherman
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3630R1C104
/ CURRENT APPLICATION NUMBER: US/10/245,621
/ CURRENT FILING DATE: 2002-09-17
/ PRIOR APPLICATION NUMBER: 10/197942
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/059114
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/063046
/ PRIOR FILING DATE: 1997-10-24
/ PRIOR APPLICATION NUMBER: 60/065027
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/086478
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/090557
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090689
/ PRIOR FILING DATE: 1998-06-25
/ Remaining Prior Application data removed - See file wrapper or PALM.
/ SEQ ID NO 90
/ LENGTH: 877
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-245-621-90

Query Match      6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVTTTCILAGCASKPTYNSTSGSGSHRTSGSG-----GLAIGSVITDSQGVNRY 73
DB 446 GAPDICVAGRLCL-S-PGCDGILGSG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGPLGY 500
QY 74 Q-----VKQGDVYSKIQRYGLNMRREIGHINNLSY-----TITYGWL----- 113
DB 501 OKILMIPAGALRLQIQL-----RPSNLTALRGPGGRSLINGNNAVDPPGSY 548
QY 114 -----TWSGDLKVERISISSGVNT-AHTPSPAV-----OSSRPVQCHPAVOKPTPPVV 164
DB 549 RAGGTATFRNRPREGKESLSABGPTTOPVDVYMIFOENRGVYQYVYISSPP---I 605
QY 165 VKKPTTPPVVQ-QP--APVAPVTEAPFATGSSGVNQF-----RYPGAIN 208
DB 606 LENPTPEPVPOLOPELIRVEPPLAPRPARTGTILQROVRIQMPAPRPHPTPLGSPA 665
QY 209 PVVRPFGTATVAGSTVTNSGMW 230
DB 666 AYWKRVGHSACSAS--CGKGVW 685

Search completed: July 8, 2003, 11:17:11
Job time : 56 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:02:47 ; Search time 41 Seconds
(without alignments)
755.007 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 1670
Sequence: 1 MVTITAINSONKPIKRLGL.....LFEFRISRNGVYDPLTVLK 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	298	17.8	297	2 S55063	conserved hypotet
2	268	16.0	314	2 T50918	hypothetical prote
3	253	15.1	242	2 A82754	lipoprotein XF0855
4	238	14.3	259	2 A65070	hypothetical prote
5	238	14.3	259	2 B91096	probable lipoprote
6	234	14.0	231	2 B83031	conserved hypotet
7	232.5	13.9	250	2 AD0871	probable lipoprote
8	219.5	13.1	405	2 E74087	lipoprotein D homo
9	219.5	13.1	454	2 E75291	probable cell wall
10	218.5	13.1	333	2 AF0407	lipoprotein lipor
11	210	12.6	373	2 AH0855	lipoprotein N1pD V
12	208.5	12.5	311	2 H82311	lipoprotein N1pD V
13	205	12.3	427	2 A13386	lipoprotein nlpD l
14	203	12.2	415	2 G81864	probable membrane
15	202.5	12.1	379	2 B55522	lipoprotein D prec
16	202.5	12.1	379	2 D91078	probable lipoprote
17	202.5	12.1	379	2 E85923	lipoprotein lipor
18	198	11.9	409	2 H81079	lipoprotein N1pD,
19	196	11.7	615	2 G87496	peptidase, M23/M37
20	188.5	11.3	537	2 AF2785	lipoprotein lipor
21	188.5	11.3	562	2 H97564	43K antigen (Afl57
22	171.5	10.3	312	2 T35413	probable secreted
23	169	10.1	223	2 F85941	conserved lipoprote
24	167	10.0	371	2 A71359	conserved hypotet
25	158.5	9.5	436	2 AH1387	cell wall binding
26	158	9.5	375	2 B75467	probable lipoprote
27	155	9.3	437	2 AB1763	cell wall binding
28	153	9.2	228	2 T36379	probable peptidase
29	152	9.1	760	2 AB2235	hypothetical prote

30	146.5	8.8	439	2 H86658	N-acetyl[muramoyl]-L
31	146.5	8.8	482	2 AG1147	P60 extracellular
32	146.5	8.8	484	2 A41487	protein P60 precu
33	139.5	8.4	465	2 AG1506	P60 extracellular
34	138.5	8.3	715	2 S76492	lipoprotein nlpD -
35	136.5	8.2	285	2 C82736	hypothetical prote
36	135	8.1	410	2 D64158	hypothetical prote
37	134	8.0	428	2 A83005	conserved hypotet
38	132	7.9	475	2 F64151	hypothetical prote
39	131	7.8	376	2 S71558	probable cell wall
40	130.5	7.8	665	2 E75461	probable cell wall
41	130.5	7.8	894	2 C86756	prophage p12 prote
42	129.5	7.8	284	2 B84118	stage II sporulati
43	129	7.7	1711	2 T31337	1,4-beta-glucanase
44	128.5	7.7	417	2 F70132	conserved hypotet
45	126.5	7.6	374	1 LYXIX	beta-lytic metallo

ALIGNMENTS

RESULT 1
S55063
conserved hypotetrical protein PA3623 [imported] - Pseudomonas aeruginosa
C/Species: Pseudomonas aeruginosa
C/Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 31-Dec-2000
C/Accession: S55063; E83193
R/Tanaka, K., Takahashi, H.
Gene 150, 81-85, 1994
A/Title: Cloning, analysis and expression of an ipos homologue gene from Pseudomonas aer
A/Reference number: S55062; MUID:95047554; PMID:7959068
A/Accession: S55063
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-297 <TAN>
A/Cross-references: EMBL:D26134; NID:G440375; PID:BA05130.1; PID:G440377
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mitsuuchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: E83193
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-297 <STO>
A/Cross-references: GB:AE004782; GB:AE004091; NID:99494772; PID:AG07011.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A/Gene: PA3623
C/Superfamily: lipoprotein D
Query Match 17.8%; Score 298; DB 2; Length 297;
Best Local Similarity 28.0%; Pred. No. 2.7e-14;
Matches 91; Conservative 50; Mismatches 106; Indels 78; Gaps 12;

QY 20 LITGVITTCILACAGK-----TNTSTSGSGSHRTSSGGLAIGQVITTSQGVN 72
DB 25 LLLGAVVCSLLAACSSSPGVKVDNNGSAPAAARTPTSG-----Q 68
QY 73 YQVKGQDVSKIAQRYGLANREIGHINNLNLSYTIYTGWLTLMSGDLKVERSSISGVN 132
DB 69 YIIRRGDTLXSLAFRCMDKALAAANGIAPTTIQVQ----- 107
QY 133 TAHTPSFVAVQ--SSRPVQGHPAVQKPTPPV--VVKKPTPTPPVQ-----QPAVAPP 184
DB 108 -----AIOFGGRASGTQ--PSVAKMTPVVAAPVATKPTPPVPAVSTSVAPADAP-A 156
QY 185 VTEAPFATGSSGMQ-----FRYPGATNPVVRRTGATVAVASTYTSNMGMTSSGRGD 238
DB 157 STTPPSSGATPVVAGVAGVGMWAPASGT--LIGRF-----ASNGSLINKGIDTLAGQIG 209
QY 239 INASNGATVIOADHNDGAS--IVIGHTNGFVSSYIIHKDAQVKTGDTVATGRIASMK 296

Db 210 VLAASGTTVYVAGSLRGELVITIKHETVSAVGNHRLVLEGQVKKGSIAMGMS 269
 QY 297 QPSGAALFEFRISRNGVYVDEPLTV 331
 Db 270 TGTDRVKLHFEIRROGKRPVDPLOYTL 294

RESULT 2

hypochemical protein nlpd [imported] - Rubrivivax gelatinosus
 C/Species: Rubrivivax gelatinosus
 C/Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
 R/Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsunura, K.; Shimada, K.
 submitted to the EMBL Data Library, November 1999
 A/Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosynth
 A/Reference number: 225270
 A/Accession: T50918
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 314
 A/Cross-references: EMBL:AB034704; PDB:BA094071.1
 A/Experimental source: strain IL14
 A/Genes: nlpd

Query Match 16.0%; Score 268; DB 2; Length 314;
 Best Local Similarity 25.6%; Pred. No. 4.2e-12;
 Matches 84; Conservative 43; Mismatches 161; Indels 40; Gaps 8;

QY 10 QNCKPRL-----GLIFGVTTCILAGCASKPTYNSTSGSGSHRTSGGLAIGSVI-- 63
 Db 8 ENSRPGTRLPAPSLILMAVAALVAGCAN-PSHRAPEVDRSGRPVASSPASAAPTAK 66
 QY 64 ----TDSQGVNRYQVQKQDVTASKIAQRYGLAMREIGHINNLSSTYITGQMLTMSGD 119
 Db 67 PPPGQENLGRFGYVAVPGDTLIRIGLENGQWMDIAFWNNLENPRILEVGVV----- 120
 QY 120 LKVERKSISGVN-----TAHTPSPVAVOSSRPVQCHPAVOKTPPPVAVVKKPTPTPV 175
 Db 121 ----RVPPGADPNGVVRPVAPARASRPASSGASAGAPAPAAAT--PAAASTAS 172
 QY 176 QCPAPVAPVTEAPFATGSGVWQCFYVPGATNPVVRRCATVAGSTYTSNGMFGSD 235
 Db 173 ASPASGSSSSSVPAARDGDDVNWTPAA-----GSVYAGFDEVRSGKLSIAGKA 223
 QY 236 GDLINSNNGVTYIOADHNMDS--IVIOHTNGFVSSYTHIKDAOYKTGDTVTRIGRIAS 293
 Db 224 GPPVVAADGRVYVAGSGIRGNTLIVGNATFTSAVAMNQTLLVKEQAVRQKIAE 283
 QY 294 MNKQPSGALFEFRISRNGVYVDEPLTV 321
 Db 284 MGASDADRQVLFELRQKRPVDPKATL 311

RESULT 3

lipoprotein Xf0855 [imported] - Xylella fastidiosa (strain 9a5c)
 C/Species: Xylella fastidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 R/Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A/Reference number: A62515; MUID:20365717; PMID:10910347
 A/Note: for a complete list of authors see reference number A59328 below
 A/Accession: A62754
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-242 <SIM>
 A/Cross-references: GB:AE003925; GB:AE003849; NID:g9105750; PDB:AAF83665.1; GSPDB:GN001
 A/Experimental source: strain 9a5c

R/Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
 Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincant, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frome
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; Laigre
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins E
 A/Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
 A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
 M.; Tenuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A/Reference number: A59328
 A/Contents: annotation
 A/Genetics:
 A/Genes: Xf0855

Query Match 15.1%; Score 253; DB 2; Length 242;
 Best Local Similarity 28.5%; Pred. No. 3.6e-11;
 Matches 87; Conservative 32; Mismatches 106; Indels 78; Gaps 9;

QY 20 LIFGVTTCILAGCASKPTYNSTSGSHRTSGGLAIGSVITTSQGVNR-YQKQK 78
 Db 10 LLSTALALNLACSTATTVPNNITSRTTA-----KPNQTVVKKQ 52
 QY 79 DTASKIAQRYGLAMREIGHINNLSSTYITGQMLTMSGD LKVERKSISGVNNTAPTS 138
 Db 53 DTLVAISRRTGVAPQDLAMNRRLPASKTIPGVGLRLVPED----- 93
 QY 139 PVAOSSRPVQCHPAVOKTPPPVAVVKKPTPTPVVQCPAPVAPVTEAPATSSGVV 198
 Db 94 -----ATASPTPQPTSPPTSPNSPTTTITTAIPANSGN 128
 QY 199 QFRYPVATNPVVRRCATVAGSTYTSNGMFGSDGLINSNNGVTYIOADHNMDSAS 258
 Db 129 WFPETGA--VVSNF-----VAGQT-TKQGVSLNGNNGTITPAANGVYVGSALIGY 180
 QY 259 --IVIOHTNGFVSSYTHIKDAOYKTGDTVTRIGRIASMNKQPSGALFEFRISRNGVYV 316
 Db 181 ELTIINKNEWMISAYGNHNRRLVNEGQIVKANQPIEM-----GMLTY-FLIRYNGKPD 234
 QY 317 PLTVL 321
 Db 235 PLTVL 239

RESULT 4

hypochemical protein b2865 - Escherichia coli (strain K-12)
 C/Species: Escherichia coli
 C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi
 Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: A65070
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-259 <BLAT>
 A/Cross-references: GB:AE000369; GB:U00096; NID:g2367168; PDB:AACT5903.1; PID:g1789228;
 A/Experimental source: strain K-12, substrain MG1655
 A/Superfamily: lipoprotein D

Query Match 14.3%; Score 238; DB 2; Length 259;
 Best Local Similarity 24.9%; Pred. No. 4.8e-10;
 Matches 78; Conservative 46; Mismatches 103; Indels 86; Gaps 8;

QY 16 KRLGLIFGVTTCILAGCASKPTYNSTSGSHRTSGGLAIGSVITTSQGVNRYQV 75
 Db 17 KSLGIYMLLSVGLIAGC-----SGSKSDPTGYSG-----SYTVT 52

QY 76 KQDVTAKIQRVGLNREIGHINNLSSTYITGQWLTMSGDLK---VRESISSGV 131
 Db 53 KRGLTVKISRTTSTSVLELRLANGISPPYITVEGQKRLT---GGAKSSSIRKSTAKSTT 110
 QY 132 NTAH-TSPVAVQSSRPVQOHPAVOKTTPPVVVKKTPPTPPVQOAPAPAPPTTEPF 190
 Db 111 KTAATVTSASVPKSSWPPVQRCWLMPTTGKVM-----PY 146
 QY 191 ATGSSGVQFRYPVGNATPVVRRFGTATVAGSTVTSNGMFPSSGRDGLINASNGTVIA 250
 Db 147 STADGG-----NKIDISAPRGFTIYAAGKAVYV 177
 QY 251 DHNMDSAS--IVIOHTNGFVSSYIHIKDAQVKTGDTVTRGRIASMKNOQPSGALLFEFRI 308
 Db 178 GNGLRGNGNLMIKHSEDEYITAYAHNDTMLVNNQGSVAGQKIATMGSTDAASVRLHFOI 237
 QY 309 SRNGVYVDPPLTVL 321
 Db 238 RYRATATDPLRYL 250

RESULT 5

B91096
 Probable lipoprotein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05093
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
 C:Accession: B91096
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gotoh, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Ref. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A:Reference number: A99629; PMID:21156231; PMID:11258796
 A:Accession: B91096
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-259 <HAV>
 A:Cross-references: GB:BA000007, PIDN:BA037161.1, PID:G13363210, GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 05093
 C:Genetics:
 A:Gene: EC03738
 C:Superfamily: lipoprotein D

Query Match 14.3%; Score 238; DB 2; Length 259;
 Best Local Similarity 24.9%; Pred. No. 4.8e-10;
 Matches 76; Conservative 46; Mismatches 103; Indels 86; Gaps 8;

QY 16 KRLLIGVITTCILAGCASKPTNSTSGSGSHRTSGSGGLAISQVITDSQVPRRYQV 75
 Db 17 KSLGIWMLSVGLLAGC-----SGSKSDPTGYSG-----SVYTV 52
 QY 76 KQDVTAKIQRVGLNREIGHINNLSSTYITGQWLTMSGDLK---VRESISSGV 131
 Db 53 KRGLTVKISRTTSTSVLELRLANGISPPYITVEGQKRLT---GGAKSSSIRKSTAKSTT 110
 QY 132 NTAH-TSPVAVQSSRPVQOHPAVOKTTPPVVVKKTPPTPPVQOAPAPAPPTTEPF 190
 Db 111 KTAATVTSASVPKSSWPPVQRCWLMPTTGKVM-----PY 146
 QY 191 ATGSSGVQFRYPVGNATPVVRRFGTATVAGSTVTSNGMFPSSGRDGLINASNGTVIA 250
 Db 147 STADGG-----NKIDISAPRGFTIYAAGKAVYV 177
 QY 251 DHNMDSAS--IVIOHTNGFVSSYIHIKDAQVKTGDTVTRGRIASMKNOQPSGALLFEFRI 308
 Db 178 GNGLRGNGNLMIKHSEDEYITAYAHNDTMLVNNQGSVAGQKIATMGSTDAASVRLHFOI 237
 QY 309 SRNGVYVDPPLTVL 321
 Db 238 RYRATATDPLRYL 250

RESULT 6

B83031

conserved hypothetical protein PA4924 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: B83031
 R:Stover, C.K.; Pham, X.Q.; Ervin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; Lim,
 . J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; PMID:20437337; PMID:10984043
 A:Accession: B83031
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-231 <STO>
 A:Cross-references: GB:AE004905; GB:AE004091, NID:G9951195; PIDN:AA08309.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4924

Query Match 14.0%; Score 234; DB 2; Length 231;
 Best Local Similarity 26.7%; Pred. No. 8e-10;
 Matches 68; Conservative 37; Mismatches 90; Indels 60; Gaps 6;

QY 69 VPNRQYQKQDVTAKIQRVGLNREIGHINNLSSTYITGQWLTMSGDLKVRERSIS 128
 Db 32 VRGEYRVKRGDTLYSIATRHGMNYKDLARANGIRPFAVKQGVVRFDRKSTYVASSR 90
 QY 129 SGVATATPSPVAVQSSRPVQOHPAVOKTTPPVVVKKTPPTPPVQOAPAPAPPTTEA 188
 Db 91 SSSNTR-----ARKPPPPPSVT----- 108
 QY 189 PFATSSGVQFRYPVGNATPVVRRFGTATVAGSTVTSNGMFPSSGRDGLINASNGTVIA 248
 Db 109 -----LKGWQPM--KEPVIRF-----SSDKLNKIRIAGTIGQPVQASLAGRV 153
 QY 249 QADHNDGAS--IVIOHTNGFVSSYIHIKDAQVKTGDTVTRGRIASMKNOQPSGALLFEF 306
 Db 154 FAVNNMGYGNLVIIQGTSTYTYANSLRLVYEGWGVGQKIATMGSTDAASVRLHFOI 213
 QY 307 RIRNGVYVDPPLTVL 321
 Db 214 EIRONGRPLDPLSL 228

RESULT 7

AD0871

probable lipoprotein [imported] - Salmonella enterica subsp. enterica serovar Typhi (str
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AD0871
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
 th, T.; Connerston, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 . S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AD0871
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-250 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD02868.1; PID:G16504121; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY3194
 C:Superfamily: lipoprotein D

Query Match 13.9%; Score 232.5; DB 2; Length 250;
 Best Local Similarity 24.4%; Pred. No. 1.1e-09;
 Matches 76; Conservative 49; Mismatches 99; Indels 87; Gaps 9;

QY 18 LGLIFGVITTCILAGCASKPTNSTSGSGSHRTSGSGGLAISQVITDSQGVPRRYQVQ 77
 Db 17 KSLGIWMLSVGLLAGC-----SGSKSDPTGYSG-----SVYTV 52

Db 11 LGIAVLCTGILLAGC-----SSNSGSGTY--SGS-----VYTVKR 44

Qy 78 GDVYSKIAORYGLNMRREIGHNINNSSTIYTGOMLTMSGDLK-----VERSISSGVT 133

Db 45 GDTLYRISRAITGTSVKELARLNGISPPYITFVGQRIRK- RGSAAKSSSTKTKNTKATKT 103

Qy 134 AHT-PSSEVAQSSRPVQOHPAVQKPTPPVYVVKKPTPPVQGPAPVAPVTEAPAT 192

Db 104 AAAPSSSSVPKSSWFPVQRCWVWPANGKVL-----PYST 139

Qy 193 GSSGVQGFRRYPVATNPVRRFGATVAGSTVTSNGMWSGGRDGLINASNACTVIOADH 252

Db 140 AEGG-----NKGDIDAGAGTTPVYAGAGKVVYVGN 170

Qy 253 NMDGAS--IYIHTNGFVSSYIHKDAQVKTGPVTRGQRIASKNQPSGALFEPRISR 310

Db 171 QLMGYGLIMIKNKEDITTAHNDTMLVANNQSVKAGQKATGTAGSTDAASVRLHFQIRY 230

Qy 311 NGVYVDELTVL 321

Db 231 RATAIDPLRYL 241

RESULT 8

lipoprotein D homolog - Haemophilus influenzae (strain Rd KW20)

C/Species: Haemophilus influenzae

C/Date: 18-Aug-1995 #sequence_rev1510 18-Aug-1995 #text_change 29-Sep-1999

C/Accession: F64087; T09410

R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.; Gocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995

A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A/Reference number: A64000; MID:95350630; PMID:7542800

A/Accession: F64087

A/Molecule type: DNA

A/Status: nucleic acid sequence not shown; translation not shown

A/Residues: 1-405 <TRIG>

A/Cross-references: GB:U32753; GB:I42023; NID:g1573701; PIDN:AAC2363.1; PID:g1573707; T A/Gene: H10706

C/Superfamily: lipoprotein D

Query Match 13.1%; Score 219.5; DB 2; Length 405;

Best Local Similarity 27.8%; Pred. No. 1.8e-08;

Matches 76; Conservative 46; Mismatches 114; Indels 37; Gaps 10;

Qy 71 NMYVQKQGTQVSKIAORYGLNMRREIGHNINNSSTIYTGOMLTMSGDLKVRERSISSG 130

Db 145 NYYKAKGDMFLINAGIDVKEALALNLSSEPNLSGOVLKISNCDIKTVTTVS-- 202

Qy 131 VNTAHIPSPVAVQSSRPVQOHPAVQKPTP-----PVVVVKKPTPPPV 175

Db 203 -----VKQAVTASATATPK--PAVTV--TPGANGTQISDGTIIIPRISSEAGTSFVVA 254

Qy 176 QQPAPVAPVTEAPATSSSGVMQFRYPVGAITPVARRFGTA--TVAGSTV---SNQMW 230

Db 255 TSSTVTSVNNANSTPTNSNVA--PI-ASN-VVMQPTSGNIIOQFSSTDSGNGKID 309

Qy 231 FSGRDGLINASNACTVIOADHNDGAS--IYIHTNGFVSSYIHKDAQVKTGDDYRTG 288

Db 310 ISGSGCAVVKAAARIVAGNALRGYGLIITIKENDPFLAVYAHNDKIIIVADQEQVYAG 369

Qy 289 QRIASMKQPSGALFEPRISRNGVYVDELTVL 321

Db 370 QDIAMGSSGTNTVTLHFIRYKSGSVDPVRYL 402

RESULT 9

E75291

probable cell wall glycyl-glycine endopeptidase - Deinococcus radiodurans (strain R1)

C/Species: Deinococcus radiodurans

C/Date: 03-Dec-1999 #sequence_rev1510 03-Dec-1999 #text_change 31-Mar-2000

C/Accession: E75291

R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma Science 286, 1571-1577, 1999

A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A/Reference number: A75250; MID:20036896; PMID:10567266

A/Accession: E75291

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-454 <WHI>

A/Cross-references: GB:A6002061; GB:A6000513; NID:g6460095; PIDN:AF11838.1; PID:g646010

A/Experimental source: strain R1

C/Genetics:

A/Gene: DR2291

A/Map position: 1

Query Match 13.1%; Score 219.5; DB 2; Length 454;

Best Local Similarity 26.1%; Pred. No. 2.1e-08;

Matches 73; Conservative 50; Mismatches 116; Indels 41; Gaps 9;

Qy 60 SOVITDSQGVNRYOVQKQGTQVSKIAORYGLNMRREIGHNINNSSTIYTGOMLTMSGD 119

Db 187 AQLAAKKPKRTTHRVEIGDTFYVAKRIGINPLADEVPRILGQTLNAGVLSVAP 246

Qy 120 LKVRERSISSGVTNHTPSPVAVQSSRP-PVQ--QHPAVQKPTPPVYVVKKPTPPPV 175

Db 247 LR-----FAVPA--PAAPAVRLVAPAPAPAPAPAPAPAPAPAPAPAPAPAP 289

Qy 176 QQPAPVAPVTEAPAT-----GSSGVQFRYPVGAITNPVRRFGTATVAGSTVY 225

Db 290 ARSPAPAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 349

Qy 226 SNGMWSGGRDGLINASNACTVIOADHND-----GASVYIHTNGFVSSYIHKDAQV 280

Db 350 HQGIDVAPAPPTVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 407

Qy 281 TGDVTRTGRIASW--KNQPSGALFEPRISRNGVYVDEL 318

Db 408 AGQLVROGERVAVGSGTGRVTPHLL-HFGLYRMDBDHNPL 446

RESULT 10

lipoprotein D

C/Species: Yersinia pestis (strain CO92)

C/Date: 02-Nov-2001 #sequence_rev1510 02-Nov-2001 #text_change 17-May-2002

C/Accession: AF0407

R/Parhilly, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Crofton, A.; Davies, R.M.; Dougan, G.; 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MID:21470413; PMID:1158360

A/Accession: AF0407

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-333 <KOR>

A/Cross-references: GB:AL590842; PIDN:CAC92586.1; PID:g15981283; GSPDB:GN00175

C/Genetics:

A/Gene: nlpD

C/Superfamily: lipoprotein D

Query Match 13.1%; Score 218.5; DB 2; Length 333;

Best Local Similarity 25.3%; Pred. No. 1.7e-08;

Matches 91; Conservative 52; Mismatches 130; Indels 87; Gaps 14;

Qy 14 PIKRLGLTFG-VITTCIIAGCASKPTVNTSGSGSHRTSGGGLAIG--SOVITDSQGV 68

Db 6 PMIRLRVAACVVALMLVGTND--NSTSAP-----ISSVGDRSGTMLSXANTDSGG 58

Qy 69 -----VP-----NRYQKQSDTVSKIAQRYGLNREIGHINNLSYITTYGQM 112
 Db 59 RIVNRSYDNI PKQSYSGNTYTVKRGDITLFYIAMI TGNDFDLAKXNNIAPRYLNVQGS 118
 Qy 113 LITMSGD---LKVREBSISSGVNTAHTPSVPAVOSSRPP---VQOHPAVQKTPPV----- 162
 Db 119 IQLNGSGGGGMIATDATTSSG-----ACPPSNIQNTTITITQNTTTTVDSQST 165
 Qy 163 -----VWKKFPPTPPVQOPAPVPVTEAPFATSSG-VMOFRYPVAGATNVV 211
 Db 166 SAYGNSCKQNVGKMLPSGSAVATTAPVAPSSSVSPASNGFVSGMRPTD----- 219
 Qy 212 RRFGTATVAGSTVTS-----NGMWFSGRDGLINASNAQTVIADHNMDCAS--IVI 261
 Db 220 -----GKTI DFSASBEGNKGIDLAGSRGCPILATAGRVVYAGNALRGYGNLII 270
 Qy 262 QHTNGFVSSYTHIKDAQVKTGDTVTRTGRIASMKRQPSGALFEFRISRNGVYVDELTVL 321
 Db 271 KHNDDYLSAVAHNDTMTLVREQOEYKAGQKIATMGSTGTSVRLHFEIRYKKGKSVNPLRYL 330

RESULT 11

AH0855
 lipoprotein NlpD precursor [imported] - *Salmonella enterica* subsp. *enterica* serovar Typh
 C/Species: *Salmonella enterica* subsp. *enterica* serovar Typh
 A/Note: this species has also been called *Salmonella typhi*
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 17-May-2002
 C/Accession: AH0855
 R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Reference number: AB0502; PMID:11677608
 A:Accession: AH0855
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-373 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD06031.1; PID:G16503998; GSPDB:GN00176
 C/Genetics:
 A:Gene: STY3050
 C:Superfamily: lipoprotein D

Query March 12.6%; Score 210; DB 2; Length 373;
 Best local similarity 23.2%; Pred. No. 8e-08;
 Matches 84; Conservative 56; Mismatches 138; Indels 84; Gaps 13;
 Qy 30 LAGCASK-----PTYNSTSGSGSHRTSG-----SGGL-----AIGS 60
 Db 23 LAGCTSSNPAPVTSVDSGSSSNTNSGMLTTPPKKATROQAPQIOPVQPTVQPMQT 82
 Qy 61 QVITDSQ-----GVV-----NRYQKQSDTVSKIAQRYGLNREIGHI 98
 Db 83 QPVTQEPVQWENGRIVNRYQGNIPKQSYTSGSTVTKKGDITLFYIAMI TGNDFDLAQR 142
 Qy 99 NNLSSTYITTYGQM-----TLMSGDLKVRERSISSGVNTAHTPSVPAVOSSRPPV- 149
 Db 143 NSISAPYSLNVQGLTVQGNASGPTTGGNATITQADAAQGVVTSANQSTVAVASQPTT 202
 Qy 150 -----QOHPAVQKTPPVVQOPAPVPVTEAPFATSSGVMQFR 201
 Db 203 YSEBSGSESAKMLPNKPKAGTATVTAFTATVSTTEENASTSAS-----ISMR 255
 Qy 202 YPVGATNVPVRRFGTATVAGSTVTSNGMFGSRDGLINASNAQTVIADHNMDCAS--I 259
 Db 256 WP--TDGTVIENFG--ASEGG---NKGIDIASGKQALIVATADGVVAVAGNALRGYGNLI 308
 Qy 260 VIOHTNGFVSSYTHIKDAQVKTGDTVTRTGRIASMKRQPSGALFEFRISRNGVYVDELTVL 319
 Db 309 IIKHNDDYLSAVAHNDTMTLVREQOEYKAGQKIATMGSTGTSVRLHFEIRYKKGKSVNPLR 368
 Qy 320 VL 321

Db 369 YL 370

RESULT 12

lipoprotein NlpD VC0533 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
 C/Species: *Vibrio cholerae*
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C/Accession: H82311
 R/Heldberg, U.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P. 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 A:Title: DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A:Reference number: AB2035; PMID:20406833; PMID:10952501
 A:Accession: H82311
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-311 <HEI>
 A:Cross-references: GB:AE004139; GB:AE003852; NID:99654953; PIDN:AAF93701.1; GSPDB:GN001.
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C/Genetics:
 A:Gene: VC0533
 A:Map position: 1

Query March 12.5%; Score 208.5; DB 2; Length 311;
 Best local similarity 25.8%; Pred. No. 8.1e-08;
 Matches 85; Conservative 48; Mismatches 138; Indels 59; Gaps 12;
 Qy 17 RLGLIFGVITTCILAGCASKRPTYNSTSGSGSHRTSGGGLAIGSQVITDSQGVPNRYQVK 76
 Db 7 RLGLL-----LCSLIFGCTA-PTAPVSGLDKDNVYKRGSGYRGSY-----YEVK 51
 Qy 77 QGDTVSKIAQRYGLNREIGHINNLSYITTYGQMLTW-----SGDLKY-- 122
 Db 52 KGDITVFIATLTQDVNDLISYNDLAPPTIHPQKIKLMPYTPPYAGTGGAATVAV 111
 Qy 123 ---RERSISSGVNTAHTPSVPAVOSSRPPVQ-----QHNAVQKTPPVVVKKPTTPP 173
 Db 112 ASSTASVAAKAATTAATVAAQTIVSKSNKVNQNGSTNSQNLTKDPTVTVQTKKEYVER 171
 Qy 174 VVOQPAPV---APPVTEAPFATSSGVMQFRYPVAGATNVPVRRFGTATVAGSTVTSNGM 229
 Db 172 VGRKNVNVNNAKAPSDK-----IAKMLP--IKGVIYKFS---AGDQ-GNKGI 216
 Qy 230 WFGSRDGLINASNAQTVIADHNMDCAS--IVIQHTNGFVSSYTHIKDAQVKTGDTVRT 287
 Db 217 DIAGORQAVATADGTVVVSGNALRGYGNLIIKNEHYLSAVAHNDQLAKEGVVQA 276
 Qy 288 GQRIASMKRQPSGALFEFRISRNGVYVDP 317
 Db 277 GQKIATMGSSGTSVRLHFEIRYKGSVNP 306

RESULT 13

lipoprotein NlpD [imported] - *Brucella melitensis* (strain 16M)
 C/Species: *Brucella melitensis*
 C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C/Accession: A13386
 R/DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Muir, C.; Los, T.; Ivanova, I.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leeseer Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*.
 A:Reference number: AD3252; PMID:11756688
 A:Accession: A13386
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-427 <KUR>
 A:Cross-references: GB:AE008917; PIDN:ALU52260.1; PID:G17983047; GSPDB:GN00190
 A:Experimental source: strain 16M
 C/Genetics:

A;Accession: G87496
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-609 <SPO>
A;Cross-references: GB:AE005673; NID:gl3423465; PIDD:AKK2371.1; GSPDB:GN00148
C;Genetics:
A;Gene: CCI1996

Query Match	11.7%	Score 196;	DB 2;	Length 609;
Best Local Similarity	23.6%	Pred. No. 1.5e-06;		
Matches 66;	Conservative 54;	Mismatches 110;	Indels 50;	Gaps 12

```

QY      73  YVQKGDVPSKIAQGYGLNWEIGHINNLSYTYTQOMTLWMSGDILKVERSTISSGVN 132
Db      347  YSVQGDITLGEIAKRFNVSVKALAEENLRLATASIKKQKIALPDGG---FRDK---GPIR 400

QY      133  TAHTSSPAVQSSRPVQOHPAVCKPTPPVVVVKKPTPTPPVQOQAPAVP---PYEA 188
Db      401  TTTTRP-----ATRPANTYARVDSASAAASPSSPVYTPSGAFAFRSAPFAVAQPIITRP 455

QY      189  PPATP-----SSGVMOFRYPGATNPVVRRCGATVAGSTVTSGMFSS 232
Db      456  P-SSQRTIETRAAPTEAEIITASGKGFAMP--RGDIISRG---VAGTGGQNDGLNIR 509

QY      233  GRDGLIANSNAGTYIQADHNND--GASIVIOHTMGCFVSYTHIMDAQVKTQDPTRTGQR 290
Db      510  APQGFPPVLSADGEIAYKAGNOVYTFGNLVLVGHAADQWTAYAHLSSTNVKRQGVKQSEQ 569

QY      291  TASM-----KQPSGALFEFR---ISRNGYNDPLTVL 321
Db      570  LGTVGATGAGVNEPQ--LHFEMRYAPTVKDKAKPFAVLVL 607

```

RESULT 20
 AF2785
 1lipoprotein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AF2785
 R:Wood, D.W.; Stethal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 et al. G.; Gillest, W.; Grant, C.; Guenther, D.; Kuyavian, T.; Levy, R.; Li, M.; McClell
 e, Karp, P.; Romero, P.; Zhang, S.
 Science 294.2317-2323. 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 R.E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577, PMID:11743193
 A:Accession: AF2785
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-537 <NUR>
 A:Cross-References: GB:AE008688, PIRN:AA42700.1, PID:q17740136, GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu1700
 A:Map position: circular chromosome

[illegible][illegible]

```

RESULT 21
H97564
43K antigen (AF157631) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C.Species: Agrobacterium tumefaciens
C.Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C.Accession: H97564
R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Gurolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294: 2323-2328, 2001
A.Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A.Reference number: A93359; PMID:11743194
A.Accession: H97564
A.Status: Preliminary
A.Molecule type: DNA
A.Restruc: 1-562 (KUB-)
A.Cross-References: GB:AE007669; PIDN:AAK87473.1; PID:G15156795; GSPDB:GN00169
C.Genetics:
A.Gene: AGR_C_3124
A.Map position: circular chromosome

```

Query Match	11.3%	Score 188.5	DB 2	Length 562
Best Local Similarity	25.1%	Pred. No. 4.8e-06		
Matches	81	Conservative	64	Mismatches 131; Indels 47; Gaps 15

QY	5	TAIINQONOKPIKRLGLIGVITTTCTTCLACASAKPTVNSGSGSHNTSGSGGIA-IGSQVI	63
DB	272	TAIANNNGQP-----TPLRAPDGNVAVLPSQAABRDKSSSEAGKTPPGGKPL	319
QY	64	TDSCGVNRYOVKQSDTVSKTAQRGLMWREIHHNNNSSSTVITGWLTLMSGDLKVR	123
DB	320	PPSG-----YKVPDSDIAKTARAGVVAALKANGI-SNESINVGQTIAI-----	366
QY	124	ERSISGVNNTATPSEFVAVOSRPRVQOHPANQKTP---PVVVYKKEPTPRPVVYQOPAP	180
DB	367	PGASIDAIKITSVPAKEA--AAAKVE--TASIKREPKAPAAATTAAPATATASVD	422
QY	181	VAPPTAEAFATSGSGVQWQFYPV-GAINPVYARPGCTAVGSGYTSNGMWSGSGDGI	239
DB	422	IEKADMASTIAESFGICIKRPFVGA---VINNGD-NVES--RNGGINISVEGPTI	476
QY	240	NASNAGTYQADHND--GASVIGHTGFGVSYTHIDAQYKTDGYRTQGRIASMKQ	297
DB	477	KRAENGVAIYAGNGIKQUGANTVLVHHDDCKVTYGNANLIDVQRQKTVQROGTIAT--SG	534
QY	298	PSGAA---LFEFRISRNGVYDP	317
DB	535	MTGSARPPVHPEVRKDALTPNF	557

RESULT 22
 t35413
 probable secreted peptidase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #ext_change 21-Jan-2000
 C:Accession: t35413
 R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 Submitted to the EMBL Data Library, March 1999
 A:Reference number: Z21577
 A:Accession: t35413
 A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA
A:Residues: 1-312
A:Cross-references: EMBL:AL049485; PIDN: CAB39706.1; GSPDB: GN00070; SCOEDB: SC6A5.22
A:Experimental source: strain A3 (2)
C:Genetics:
A:Gene: SCOEDB: SC6A5.22
C:Superfamily: Lipoprotein D

Query Match 10.3%; Score 171.5; DB 2; Length 312;
Best Local Similarity 23.3%; Pred. No. 3.8e-05;
Matches 79; Conservative 43; Mismatches 118; Indels 99; Gaps 15;

```
Qy 46 GSHR-----TSGSGGLAIGSVITDS-----QGVF-----70
Dy 5 GKRRRTAMRLRTTAVAGTGGAALPLMGAAANAPASVSEQAVQSPASAKKAA 64
Qy 71 -----NRQYQGDPTVSKIAQRYGLMREIGHINNNSSITTYTGQWLTMSGDLKRE 124
Dy 65 EKNSDSRTYTKSGDYSLKDEBDVD-----GGMKKLYADN---R 102
Qy 125 RSISGVTATHTSPVAV--QSSRPVQHPAVOKPPTPVVVKKPTPTPVVQOPAPVA 182
Dy 103 EAVSDPELHPGLKLSIDGAAKAPASASQKPA-----QKAEKPAEKTAAP 155
Qy 183 PPTVAPATGSSGVMQFRYPVGAATNPVRRFGTATVAGSTV-----TSNGMFSG--- 233
Dy 156 QKSTDAEKASSSDQSGQ---STGTTSGV-----TSPVAGTGVGTPYHQSSGMSGHTG 207
Qy 234 -----RQGDILNANAGTIVQAD--HNMDSIVIQHTNGFVSSYIHKDAQVKTGDTYRT 287
Dy 208 TDFVVPPTGSLKAVAGAGTAVSAGWGAAGVQVQLADGHVAYAHLSLSVSAGOSVTA 267
Qy 288 GQRI--ASMKNQPSGAL--FEFRISRN--GVVVDPLTVLK 322
Dy 268 GQVGLSGATGNTGPHLHFEIRITTPDYGSDIDPLAFK 306
```

RESULT 23

Probable lipoprotein Z4203 [imported] - *Escherichia coli* (strain O157:H7, substrain EDLS F85941)
C:Species: *Escherichia coli*
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Dec-2001
C:Accession: F85941
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001.
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85941
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-223 <STO>
A:Cross-references: GB:AE005174; NID:912517383; PIDN: AAG57994.1; GSPDB: GN00145; UMG: Z42
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z4203
C:Superfamily: Lipoprotein D

Query Match 10.1%; Score 169; DB 2; Length 223;
Best Local Similarity 24.0%; Pred. No. 3.8e-05;
Matches 63; Conservative 35; Mismatches 79; Indels 86; Gaps 8;

```
Qy 29 ILACASKPTNSTRSGSHRTSSGGLAIGSVITDSQVPMNVQYQGDPTVSKIAQRY 88
Dy 9 LLAGC-----SGSKSDPTGYSG-----SVYTVRRGDTLIRISRTT 44
Qy 89 GLMREIGHINNNSSYTYTGQWLTMSGDLK-----VREKSISSGVNTAH--TSPVAVQ 143
Dy 45 GTSVVELARLNGISPPYITLVGQKTLK--GSAKSSSRKSTAKSTTKHSAVTSSAVPK 102
Qy 144 SSRPPVQHPAVVQKPTPTPVVVKKPTPTPVVQOPAPVAPVTAEPATGSSGVMQFRYP 203
Dy 103 SSMPVQGRCLWLTPTGKVM-----PYSTADGG----- 131
```

```
Qy 204 VGATNPVRRFGTATVAGSTVTSNGMWFSGRDDLINASAGTIVQADHMDGAS--IV 261
Dy 132 -----KKGIDISAPRGPIPIYAAAGAKVYVGNQLRGIGNLIMI 169
```

```
Qy 262 QHTNGFVSSYIHKDAQVKTGDT 284
Dy 170 KHSEDTITAYAHNDTMLVNNQGS 192
```

RESULT 24

A:1359
convered hypothetical protein TP0155 - *Syphilis spirochete*
C:Species: *Treponema pallidum* subsp. *pallidum* (*Syphilis spirochete*)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: A71359
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Ueterbach, T.; McDo
ney, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of *Treponema pallidum*, the *Syphilis spirochete*.
A:Reference number: A71250; MUID:9832770; PMID:9655876
A:Accession: A71359
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-371 <COL>
A:Cross-references: GB:AE001200; GB:AE000520; NID:93322419; PIDN: AAC65145.1; PID:9332242;
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0155

Query Match 10.0%; Score 167; DB 2; Length 371;
Best Local Similarity 24.9%; Pred. No. 0.0001;
Matches 72; Conservative 47; Mismatches 122; Indels 48; Gaps 11;

```
Qy 73 YQVQGDPTVSKIAQRYGLMREIGHINNNSSYTYTGQWLTMSGD---LKYREKSISS 129
Dy 87 YEVEGDVGRVRIQRYDVISQDAIISLKLSTALQVGLKIPSVGILYTVNKGDTFS 146
Qy 130 GVTATHTSPVAVVQSSPPVQHPAVOKPPTPVV--VKKP-----TPTP----- 172
Dy 147 SIAAH-----QSLLELVILNTPSSSKESPVRILTVPFYNASARESCVPPFPSSAKQ 201
Qy 173 -----PVQ--QPAVAPVTEAPATSSGVMQ-----FRYPGATNPVRRFG-- 215
Dy 202 WRENTSDAVQIOPA---RVLFPEAHLSARALQELNGBDLFRAPLRSTYVSSRYGMR 257
Qy 216 TATVAGSTVTSNGMWFSGRDDLINASAGTIVQADH--MDGASIVIQHTNGFVSSYIHI 274
Dy 258 SDPFTGARSFHNGLDMVSRRGITVYSALGIVATVGSANVGNVLYIGHAGVDTLXGHL 317
Qy 275 KDAQVKTGDTVTRTGQRIASW--KNQPSGALFEFRISRNQVYVDDPLTVLK 322
Dy 318 QTVLVSAGTAVTSATKIGLIGKTRGSRGPHLFTIYNAGSAINPTSLR 366
```

RESULT 25

AH1387
cell wall binding proteins homolog lmo2504 [imported] - *Listeria monocytogenes* (strain Ek
C:Species: *Listeria monocytogenes*
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH1387
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entlian, K.D.; Feihl, H.,
D.; Jones, L.M.; Karet, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueder, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of *Listeria species*.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1387
A:Status: preliminary
A:Molecule type: DNA

A/Residues: 1-436 <GLA>
A/Cross-references: GB:NC_003210; PIDN:CAD00582.1; PID:g16411592; GSPDB:GN00177
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: lmo2504

Query Match
Best Local Similarity 24.8%; Pred. No. 0.00051;
Matches 66; Conservative 33; Mismatches 100; Indels 67; Gaps 11;

QY 76 KQDVTYSKIAQRVGLNWRERIGHNNLSYTYTGQWLTMSGDLKVRERISGGVNTAH 135
DB 210 EKNDLVMLANKKDLT-----KSEQTLLAB-----QGLTDEKRLASINAGEK 254
QY 136 TSPVAVOSSRPVVOQHPRVVKKPTTPPVVVOQAPVAPVTEAPFATGSS 195
DB 255 AKQEAATKAAEKRMQEAALASAKSAVVK-----QPSSSNEATE-----IVSS 300
QY 196 GVMOPRYVGA-----TPVVRFR-----GTAIVAGSTVTSNGMFSGRDGLINAS 242
DB 301 GGGQPIKPAASGLTSGFSERTNPVTKESHKGODIAGGGTIT-----VSA 347
QY 243 NAGTVIQADHNMDS-----IVIOHTNGFVSSYHIKDAQVK--TGDTVTRGORIAS 293
DB 348 ASGTIVVFSFGASGSGFGGYGVVVIDHNGFQTLVGHMRASLKVTVGQGVSGQPIGI 407
QY 294 M--KNQPSGALFEFRISRNQVYDP 317
DB 408 MGTSGQSTGQHL-HFEIHKNGIPIVD 432

RESULT 26
F75467
probable lipoprotein - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C/Accession: F75467
R.White, O.; Eichen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036896; PMID:10567266
A/Accession: F75467
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-375 <WHI>
A/Cross-references: GB:AE001939; GB:AE000513; NID:G6458563; PIDN:AAF10427.1; PID:G645856
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR0648
A/Map position: 1

Query Match
Best Local Similarity 9.5%; Score 158; DB 2; Length 375;
Matches 74; Conservative 36; Mismatches 105; Indels 126; Gaps 12;

QY 68 GPNRYVYKQDVTYSKIAQRVGLNWRERIGHNNLSYTYTGQWLTMSGDLKVRERIS 127
DB 14 GVAAYVYKKDDTLYSLARSGGLTVDALMRNLNGI-STPELRVGVYIKL-----P 61
QY 128 SSGVNTAHTSPVAVOSSRP-PVOQHPRVVKKPT--PPV-----VVKKP--TPRP----- 173
DB 62 GEGATATPATPAPAPPTATPAPAPAPAVFTLPTVATVAVNLKLPVITITAPALMK 121
QY 174 -----VVOQAPVAPVPTV-----APFATGSSGVMQFPRPVATPVV- 211
DB 122 GDGFALRISGRQANVTVVRPSELGEDVVRQNELRLMSGGYIYVGRVVLGKTIPVY 181
QY 212 -----RR----- 213
DB 182 EVRLGDPVVRGQIPVQDLGQRVGHNLNPSISKVLQDPAREADDAVEQAYQKRTQQQW 241

QY 214 --PGTAVAGSTVTSN-----GVMFSGRDGLINASAGTVIQAD--HN 253
DB 242 QPFAASLAAAGKATSSSFGQPRVTVAGKVAHYGDIYARSGTAVLINDGTIVYAGNYP 301
QY 254 MDGASIVIOHTNGFVSSYHIKDAQVKGDVTRTGORIASM 294
DB 302 VKGSLVIVHNGAGVSVLYHNSKVTAKVGGQIKRQKQVGEV 342

RESULT 27
AB1763
Cell wall binding protein homolog lln2647 [imported] - Listeria innocua (strain Clp1126;
C/Species: Listeria innocua
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C/Accession: AB1763
R.Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Feihl, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A/Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A/Title: Comparative genomics of Listeria species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AB1763
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-437 <GLA>
A/Cross-references: GB:AL592022; PIDN:CAQ97874.1; PID:g164115184; GSPDB:GN00178
A/Experimental source: strain Clp11262
C/Genetics:
A/Gene: lln2647

Query Match
Best Local Similarity 9.3%; Score 155; DB 2; Length 437;
Matches 62; Conservative 35; Mismatches 103; Indels 66; Gaps 10;

QY 76 KQDVTYSKIAQRVGLNWRERIGHNNLSYTYTGQWLTMSGDLKVRERISGGVNTAH 135
DB 210 EKNDLVMLANKKDLT-----KSEQTLLAB-----QGLTDEKRLASINAGEK 254
QY 136 TSPVAVOSSRPVVOQHPRVVKKPTTPPVVVOQAPVAPVTEAPFATGSS 195
DB 255 AKQEAATKAAEKRMQEAALASAKNAAT-----VAAQPSVTSAGTATPTVSS 301
QY 196 GVMOPRYVGA-----TPVVRFR-----GTAIVAGSTVTSNGMFSGRDGLINAS 242
DB 302 GGGQPIKPAASGLTSGFSERTNPVTKESHKGODIAGGGTIT-----VSA 348
QY 243 NAGTVIQADHNMDS-----IVIOHTNGFVSSYHIKDAQVK--TGDTVTRGORIAS 293
DB 349 ASGRVVFSGFATGSGFGGYGVVVKIDHNGFQTLVHMRASLKVTVGQGVSGQPIGI 408
QY 294 M--KNQPSGALFEFRISRNQVYDP 317
DB 409 MGTSGQSTGQHL-HFEIHKNGIPIVD 432

RESULT 28
T36379
probable peptidase - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: T36379
R.Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, April 1999
A/Reference number: Z2153
A/Accession: T36379
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-228 <OLI>
A/Cross-references: EMBL:AL049628; PIDN:CA840868.1; GSPDB:GN00070; SCOEDB:SCB94.19c
A/Experimental source: strain A3(2)
C/Genetics:

A:Gene: SCOEDB:SC94.19c

Query Match 9.2%; Score 153; DB 2; Length 228;

Best Local Similarity 37.7%; Pred. No. 0.00055;

Matches 46; Conservative 16; Mismatches 42; Indels 18; Gaps 6;

QY 219 VAGSTVTSNGM---FSGRD-----GDLINASNAGTVIQADHN-----MDGASIVYQHT 264

DB 101 LBSAFNONGMNAHKISGDDFAVPIGTNNVVAHAGTIVKAGNGAGDGPAYGNAYIKHG 160

QY 265 NGPVSSYIHKDAQVTCGTTCGTGQRIASMKN--QPSGAL--FEFRISRN--GYVDPPLTV 320

DB 161 NCTYSQYAHLSRINVIKGIQIVKIGOSIAGSGNNGSGPLHPEIRITTPYGSADVPAV 220

QY 321 LK 322

DB 221 LR 222

RESULT 29

AB2225 hypothetical protein alr3353 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AB2225

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iiziguchi, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AB2225

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-760 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA075052.1; PID:g17132448; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3353

Query Match 9.1%; Score 152; DB 2; Length 760;

Best Local Similarity 20.1%; Pred. No. 0.0031;

Matches 95; Conservative 41; Mismatches 112; Indels 224; Gaps 18;

QY 61 QVITD-SQGVPRVYQKQDVTSKIAQRYGLMREIGHNNINNSSTIYTGWLTLMSGD 119

DB 294 QPLTETASTANTVEKPEDTILALASRYNTSVAEIVKVNLSNPOLKISQGLIIPS-- 351

QY 130 LKVRERSISSGVNTATPS-----PVAVOSRP--PVQGHFAVQK-- 157

DB 352 -----AIDSSSTIAQAPALISSNRVQGTPTVPFPVDTARVNSLVAQPPSIANNND 405

QY 158 -----PTP-----PV-----VVKKP----- 168

DB 406 SVPITVPTPGADSETPVDITVPLESASAPAEIQGVGNVPIPKAFIEIOPQPGKRAAR 465

QY 169 -----TPPPVQO----- 177

DB 466 AKGDRLSIQAEIQRLOAKYRDOQGTNTVPVANNNTALPIPVTSNNPAVTRPISR 525

QY 178 -----PAPVAPVTEAPFATGSSGVNQFR--YVQAT-----NPVRRFGTATV 219

DB 526 QOEIVAPVIAVPTPLPNDNNHVPKQFRATLPPNMLNPEFLPNNAPQKPSRVATPPV 585

QY 220 -----AGSTY-----TSNG 228

DB 586 RLNAESLGRMRGTTVSPKPLPLAANDQYLPOAIDESVPPSPDSTAFITPAKGVLTSGV 645

QY 229 MWFSGR-----DGLINASNAGTVIQADHNMDGAS--IVIOHTGFPVSYTHID 276

DB 646 GRRKGRMHRGIIIAAPVGTVPALADGVVEKSGMNGGYNGLVDINHDPGSLTRVHNSR 705

QY 277 AQVKTGDTVTRTGORIASMKOPSALF-----EFRISRNGV--YVDPPLTVL 321

DB 706 LTVQAGQVYRQGGQIRAM-----GSTGFGSTGPHTHFEIHKTGKGAINPIAML 752

RESULT 30

H86658 N-acylglutamate-L-alanine amidase (EC 3.5.1.28) [imported] - Lactococcus lactis subsp.

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

C:Accession: H86658

R:Boletín, A.; Kinner, P.; Mager, S.; Jallón, O.; Malarme, K.; Weisenbach, J.; Ehrli

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp

A:Reference number: AB6625; MUID:21235186; PMID:11337471

A:Accession: H86658

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-439 <STO>

A:Cross-references: GB:AE005176; PID:g12723133; PID:AAK04370.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: acma

C:Keywords: hydrolase

Query Match 8.8%; Score 146.5; DB 2; Length 439;

Best Local Similarity 23.6%; Pred. No. 0.0038;

Matches 69; Conservative 39; Mismatches 107; Indels 77; Gaps 10;

QY 26 TFCIAGCASKPTYNST-----SGSGHRTSGSGGLAIGSQVITDSQGVPR 72

DB 184 TALTGKYATDPYVGASLRIRISQVNLTRFDGASAGTSNGC-STATNNNSNTSSTT 242

QY 73 YQKQDVTYSKIAQRYGLMREIGHNNINNSSTIYTGWLTLMSGD LKVRERSISSGVN 132

DB 243 YTVKSGDTLMGTSQKIGISVAQIQSANLKSST-VIYIGQLVLTSSSSSNTNSSTSGN 301

QY 133 THTPSPVAVOSRPPVQGHFAVQKPTPPVVKKPTPPVQVQAPVAPVPT---EA 188

DB 302 SAGTTPTT-----STPKAPASQTIHKKVSGDTLWGLSLV 336

QY 189 PRATGSSGVNQFRYPVQATNPVRRFGTATVAGSTVTSNGMFSRGDGLINASNAGTVI 248

DB 337 KYKTTIAQLKSNHNLNSDTIFIGQNLIVSQSAGSSSSSTG-----SSSAST-- 382

QY 249 QADHNMDGASIVYQHTNGFVSVYHKDAQVKTGDPV-----RTGORIASMK 295

DB 383 SSTNSASAS-----NTSIH---KVKAGDTLMGLSQKSGSPIASIK 420

RESULT 31

AG1147 P60 extracellular protein, invasion associated protein Iap [imported] - Listeria monocytogenes

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AG1147

R:Glaser, P.; Frangoul, L.; Buchleser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesuguet, O.; Entian, K.D.; Feihl, H.; Jones, L.M.; Karsc, U. Science 294, 849-852, 2001

A:Authors: Krefel, J.; Kuhn, M.; Kunat, F.; Kurapat, G.; Madueno, E.; Maltournam, A.; Matok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A. Title: Comparative genomics of Listeria species. A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1147

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-482 <GLA>

A:Cross-references: GB:NC 003210; PIDN:CAC98661.1; PID:g16409958; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: iap

Query Match 8.8%; Score 146.5; DB 2; Length 482;

Best Local Similarity 20.6%; Pred. No. 0.0043;
 Matches 81; Conservative 46; Mismatches 121; Indels 145; Gaps 14;

QY 25 ITTCILAGCASKPTNSTSGSHRTSGGGLA--ISQVITD----- 65
 DB 100 IITSIKGG--TKYVETTESNGMHKITYNDDKGTGVNGKYLTDKAVSTPAAPQEVKETT 157
 QY 66 --SQGV-----NRVQKQDPTVSKIAQRYG 89
 DB 158 TTQQAAPAAETKEVKTQTQATTAPKVAETKEPVPVDDQATTHAVSGDITMALSVKYG 217
 QY 90 LNMREIGHINNLSSTYITYGWLTLMGDLKVERSISSGVNTAHTPSPVAVQSRPVP 149
 DB 218 VSVODIMSNMNLSS--SIYVGQKLA-----KQTANTATPKAEVKT----- 257
 QY 150 QQHPAYOKPTPPV-----VKKPTPPVQOAPAPVAPPTVTEAPFATGSGVWQ 199
 DB 258 -EAPAAEKQAPVVKENTNTNTATTEKETATQ--QQTAPKAPTEAKPAPAPSTNTNA 313
 QY 200 FRYPVGA-----TNPVRRFGATVAGSTVTSNGMWFSGRDGDLINSMNGTVIQ 249
 DB 314 NKTNT 373
 QY 250 ADHNM-----GASIVIGHTNGF--VSSYIHKDAQVYTGPT 284
 DB 374 AQKHGKAVSWGNGGPTTDCSGYTKVYFAKAGISLPTSGAQYASTRTISBSQAKPGDL 433
 QY 285 VRTGQRIASMKNOPSGALFEF--RISNGVYV 315
 DB 434 V-----FFDYGSGISHVGIV 449

RESULT 32

A1487
 protein p60 precursor - *Listeria monocytogenes*
 N/Alternate names: invasion-associated protein
 C/Species: *Listeria monocytogenes*
 C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 13-Sep-1996
 A/Accession: A1487, B41487
 R/Kochler, S.; Leimeister-Maehter, M.; Chakraborty, T.; Lottspeich, F.; Goebel, W.
 Infect. Immun. 58, 1943-1950, 1990
 A/Title: The gene coding for protein p60 of *Listeria monocytogenes* and its use as a spec
 A/Reference number: A1487; PMID:90256283; PMID:2111287
 A/Accession: A1487
 A/Molecule type: DNA
 A/Residues: 1-484 <KOE>
 A/Cross-references: GB:X52268
 A/Accession: B41487
 A/Molecule type: protein
 A/Residues: 28-49 <KO2>
 C/Genetics:
 A/Gene: *iap*
 F11-27/Domain: signal sequence #status predicted <SIG>
 F128-484/Product: protein 60 #status predicted <MAT>

Query Match 8.8%; Score 146.5; DB 2; Length 484;
 Best Local Similarity 20.6%; Pred. No. 0.0043;
 Matches 81; Conservative 46; Mismatches 121; Indels 145; Gaps 14;

QY 25 ITTCILAGCASKPTNSTSGSHRTSGGGLA--ISQVITD----- 65
 DB 102 IITSIKGG--TKYVETTESNGMHKITYNDDKGTGVNGKYLTDKAVSTPAAPQEVKETT 159
 QY 66 --SQGV-----NRVQKQDPTVSKIAQRYG 89
 DB 160 TTQQAAPAAETKEVKTQTQATTAPKVAETKEPVPVDDQATTHAVSGDITMALSVKYG 219
 QY 90 LNMREIGHINNLSSTYITYGWLTLMGDLKVERSISSGVNTAHTPSPVAVQSRPVP 149
 DB 220 VSVODIMSNMNLSS--SIYVGQKLA-----KQTANTATPKAEVKT----- 259
 QY 150 QQHPAYOKPTPPV-----VKKPTPPVQOAPAPVAPPTVTEAPFATGSGVWQ 199

DB 260 -EAPAAEKQAPVVKENTNTNTATTEKETATQ--QQTAPKAPTEAKPAPAPSTNTNA 315
 QY 200 FRYPVGA-----TNPVRRFGATVAGSTVTSNGMWFSGRDGDLINSMNGTVIQ 249
 DB 316 NKTNT 375
 QY 250 ADHNM-----GASIVIGHTNGF--VSSYIHKDAQVYTGPT 284
 DB 376 AQKHGKAVSWGNGGPTTDCSGYTKVYFAKAGISLPTSGAQYASTRTISBSQAKPGDL 435
 QY 285 VRTGQRIASMKNOPSGALFEF--RISNGVYV 315
 DB 436 V-----FFDYGSGISHVGIV 451

RESULT 33

AG1506
 p60 extracellular protein, invasion associated protein *iap* [imported] - *Listeria innocua*
 C/Species: *Listeria innocua*
 C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C/Accession: AG1506
 R./Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker
 .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fsihi, H.
 Science 294, 849-852, 2001
 D.; Jones, L.M.; Karet, U.
 A/Authors: Kreft, U.; Kuhn, M.; Kuntz, F.; Karapakt, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schlueter, T.; Simoes, N.; Tietzel, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
 A./Title: Comparative genomics of *Listeria species*
 A/Reference number: AB1077; PMID:11679669
 A/Accession: AG1506
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-465 <GLA>
 A/Cross-references: GB:AL592022; PIDN:CAC95823.1; PID:G16413031; GSPDB:GN00178
 A/Experimental source: strain Clp11262
 C/Genetics:
 A/Gene: *iap*

Query Match 8.4%; Score 139.5; DB 2; Length 465;
 Best Local Similarity 21.4%; Pred. No. 0.013;
 Matches 75; Conservative 51; Mismatches 143; Indels 87; Gaps 14;

QY 25 ITTCILAGCASKPTNSTSGSHRTSGGGLA--ISQVITD----- 65
 DB 99 IITSIKGG--TKYVETTESNGMHKITYNDDKGTGVNGKYLTDKAVSTPAAPQEVKETT 156
 QY 66 --SQGV-----NRVQKQDPTVSKIAQRYG 90
 DB 157 QQVPATEAKTEVKOPTTQQTAPAPKAEKTEAPVDDQATTHAVSGDITMALSVKYG 216
 QY 91 NLMREIGHINNLSSTYITYGWLTLMGDLKVERSISSGVNTAHTPSPVAVQSRPVP 150
 DB 217 SVODIMSNMNLSS--SIYVGQKLA-----IKOPTVAPPAET-KTQAPAAEKQTPAVK 269
 QY 151 QHPAVOKPTPPVVKKPTPPVQOAPAPVAPPTVTEAPFATGSGVWQFRYVVGATNP 210
 DB 270 ENSNANTAT-----TEKETATE-----QQTTPAQTQAKPAPAPSTNTNTNTNTNTNS 322
 QY 211 VRRFGATVAGSTVTSNGMWFSGRDGDLINSMNGTVIAD-----HNMDSAGIVIGHT 264
 DB 323 TPKSKTNT 377
 QY 265 NGFVSTYIHKDA--QVKTGDTVTRGQRIASMKNOPSGALFEF--RISNGVYV 315
 DB 378 SGF-TKYVFAKSGISLPTSGAQYASTRTISBSQAKPGDLVFFDYGSGISHVGIV 432

RESULT 34

S76492
 lipoprotein nlpD - *Synechocystis* sp. (strain PCC 6803)
 N/Alternate names: protein nlpD993
 C/Species: *Synechocystis* sp.
 A/Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C/Accession: S76492
 R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimoto, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-116, 1996
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 sp.
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S76492
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-715 <KAN>
 A/Cross-references: EMBL:D90915; GB:AB001339; NID:9163604; PID:BA18621.1; PID:d101935
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C/Genetics:
 A/Gene: nlpD

Query Match 8.3%; Score 138.5; DB 2; Length 715;
 Best Local Similarity 22.7%; Pred. No. 0.027;
 Matches 76; Conservative 44; Mismatches 128; Indels 87; Gaps 12;

QY YVKGQDVTSKIAQRVGLNWRIGHINNLSYITYGOMLTMSGD-----KYRERSI 127
 DB YVKGQDVTSKIAQRVGLNWRIGHINNLSYITYGOMLTMSGD-----KYRERSI 127
 QY 128 SSGVNTAHTPSPVAVQSRPPVQV----- 152
 DB 438 PPSFSTGPTASTIYARAQAEPBAQVQVQLRAEVQLNQTQPIRTPMVRSALTVAQVNN 497
 QY 153 ---PAVQKPTPPVVVVKKPTPPPVVQGPAPVAPVTEAPAT--GSSGVN----- 198
 DB 498 EPVDMQARSPQALPAKFNQPRQDAQLQROQSPQAQSGFSTVGSQGIYAAPSPVQ 557
 QY 199 ---QRRYV-----GATNPVVR-----GATVAGSTVSN-----GMFSSGD- 235
 DB 558 YNDTITLIVIQEVSPELFGSLTDPFRRSPGQFNGITPAKGVFTSGFPPRRMRGRDI 617
 QY 236 ---GDLINASNAGVIOADHNMDS--IYIQTNGFVSSYIHKDAQVKTGDTVRTQ 289
 DB 618 AAPGTPTMAAASGVVSGNNSGFGNLVKIRHDSVTYAHNNRLLVARGEVESQ 677
 QY 290 RIASM-KNPSGAL-FEPRISRNGVYDPLTVL 321
 DB 678 QIAEWGSGRSTGPHLFEIRVGTNA-VNPAVL 711

RESULT 35
 C82736
 hypothetical protein XF0987 (imported) - Xylella faecidiosa (strain 9a5c)
 C/Species: Xylella faecidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C/Accession: C82736
 R/anonymous: The Xylella faecidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A/Title: The genome sequence of the plant pathogen Xylella faecidiosa.
 A/Reference number: A62515; MUID:20365717; PMID:10910347
 A/Note: for a complete list of authors see reference number A59328 below
 A/Accession: C82736
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-285 <SIM>
 A/Cross-references: GB:AB003937; GB:AE003849; NID:99105920; PID:AAF83797.1; GSPDB:GN001
 A/Experimental source: strain 9a5c
 R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H
 ae-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; From
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laig
 chado, M.A.; Medeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E
 A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tshako, M.H.; Valiada, H.; Van Sluys, M.A.; Verjovskij-Almeida, S.; Vettore, A.L.; Z
 A/Reference number: A59328
 A/Contents: annotation
 C/Genetics:
 A/Gene: XF0987

Query Match 8.2%; Score 136.5; DB 2; Length 285;
 Best Local Similarity 22.8%; Pred. No. 0.011;
 Matches 76; Conservative 49; Mismatches 126; Indels 83; Gaps 17;

QY 15 IKRLGIFGVITTCILAC--ASKPTNSTSGSGSHRTSGSGGLAISQVITDSQVNR 72
 DB 6 VNRL-LALGMLVTLTLPITAYAKKKTOTRSEFPNVGALVIGT-----VPAG 56
 QY 73 YVKGQDVTSKIAQRVGLNWRIGHINNLSYITYGOMLTMSGD 119
 DB 57 SRVEYSGQVLRVDKGVVFGVGNNGIPILRQSEGVTKQVNTVTRQW----- 109
 QY 120 LKVERISSGVNTAHTPSPVAVQSRPPVQGPAPVAPVTEAPAT--GSSGVN----- 152
 DB 110 -----PLEVNGVPKTVN-----PELTAAR-----ITRBA 137
 QY 180 PVADPTVTEAPATG--SSGVNQFRYPVGAATNPVVRFGTATVAGSTVTSNGMFSGD-- 235
 DB 138 ---QITDARQNRDNRSDFAQAFIWPV--QGRISGRFSARIYNGTAAAG--HSGMDIA 189
 QY 236 ---GDLINASNAGVIOA--DHNMDSYIYQHTNGFVSSYIHKDAQVKTGDTVRTQ 290
 DB 190 TATPTPKAPAPAGVITPAADPLVLTGTVLDHGAQVSNFHLISRIDVKGVDHVGQV 249
 QY 291 IASM-KNPSGAL-FEPRISRNGVYDPLTVL 322
 DB 250 IGAVGATGRATGPHL-HMGWMPVVRIDPLVLE 282

RESULT 36
 D64158
 hypothetical protein H10756 - Haemophilus influenzae (strain Rd KW20)
 C/Species: Haemophilus influenzae
 C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
 C/Accession: D64158
 R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
 J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter, J
 A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A/Reference number: A64000; MUID:95350630; PMID:7542800
 A/Accession: D64158
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-410 <TIGR>
 A/Cross-references: GB:U32760; GB:L42023; NID:91573764; PID:AA022415.1; PID:91573765; T
 A/Note: best homolog was a hypothetical protein from Escherichia coli

Query Match 8.1%; Score 135; DB 2; Length 410;
 Best Local Similarity 25.0%; Pred. No. 0.023;
 Matches 63; Conservative 36; Mismatches 111; Indels 42; Gaps 10;

QY 74 YVKGQDVTSKIAQRVGLNWRIGHINNLSYITYGOMLTMSGD-----KYRERSI 127
 DB 188 QXKQQAALQKQAEHQ-----STLNLNLKMLLDQKNT-----LKAEQALROEIOR 236
 QY 134 AHTPSPVAVQSRPPV-QQHPAVOKPTPPVVVVKKPTPPPVVQGPAPVAPVTEAPAT 192
 DB 237 AEQAPARQERERREALQROKAEKRT-----SKPYOPTVGEROLLN 278
 QY 193 GSSGV-----MQRPVGATATPVVRRTATVAGSTVTSNGMWSGRCGLINSNAGTIV 248
 DB 279 STSLGAQAKQVSLPVG--ILHTFG--SIQAGEVWKMVIGASGTPVKALAAARVI 334
 QY 249 QADH-NMDGASIVYQHTNGFVSSYIHKDAQVKTGDTVRTQRIASMKN--QPSGALFE 305

Db 335 LAGYLNGVGYVTKKGETDLSTLYGPNQAVSVKQGLVAGQVIAQVGTGEISRSALY 393
QY 306 FRISRGVYVDP 317
Db 394 FGISRGKGPVNP 405

RESULT 37
A83005
conserved hypothetical protein PA5133 (imported) - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: A83005
R/Author: C.K.; Phan, X.Q.; Edwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lim,
.; Loay, S.; Olson, M.V.
N/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
gen.
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: A83005
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-428 <STO>
A/Cross-references: GB:AE004926; GB:AE004091; NID:99951424; PIDN:AAG08518.1; GSPDB:GNO01
A/Experimental source: strain PA01
C/Genetics:
A/Genes: PA5133

Query Match 8.0%; Score 134; DB 2; Length 428;
Best Local Similarity 28.2%; Pred. No. 0.029;
Matches 53; Conservative 28; Mismatches 79; Indels 28; Gaps 9;
QY 140 VAVQSRPVQOHPAVQKPTPPVYVKKPTPPVYVQGPAPVAPVTEAPVATGSGVWQ 199
Db 255 LAERRRARRQQAAPRVTSP-----REBAPGVLVS-----TGAVYG-GARGAR 300
QY 200 FRYPVATNPVYRRPCTATVAGSTVNSGMWFGSGRDGLINASNGTIVQADHNDGAS- 258
Db 301 GKLTPVNGVNVARFSGRDGDDPRAKMDVLISASGSTRVAVHGRVVFADW-LRGAQL 359
QY 259 -IVIOHTNGFVSSYH-----IKDAQVKTGDTVPTGRIASM--KNPSCGALFEPRISN 311
Db 360 LVLDHGCGTSLYHGNOSILKDA---GDTVAAGPPIATVGTSGGSSPAAV-PAIRKQ 414
QY 312 GYVVDPLT 319
Db 415 GRPADPPT 422

RESULT 38
F64151
hypothetical protein HI0409 - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C/Accession: F64151
R/Author: R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman, J
.; D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: F64151
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1475 <TRF>
A/Cross-references: GB:U32724; GB:U42023; NID:91573379; PIDN:AAC22068.1; PID:91573382; T
A/Note: Best homolog was a hypothetical protein from Escherichia coli
A/Superfamily: Hypothetical protein HI0409

Query Match 7.9%; Score 132; DB 2; Length 475;
Best Local Similarity 23.0%; Pred. No. 0.047;

Matches 67; Conservative 37; Mismatches 83; Indels 104; Gaps 12;
QY 72 RYQVQGD---TWKIAQRYLANMREIGHINLT-----NSSYT-----YTQQ 111
Db 217 RKEVAKGELQNSLNSLSREKGLDTRQISOLSNALQWQSLKLLKKGQFALVSREYIGD 276
QY 112 WLTLSGDLKVRERSISGVTATPSPVAVOS-----SRPPVQOHPA 154
Db 277 KLT---GGQVNEALRISGGKNY-----AVQANGRYNQOGETLKKGFARVYLQROAR 328
QY 155 VOKPTPPVYVKKPTPPVYVQGPAPVAPVTEAPVATGSGVWQFRYPVATNPVRRF 214
Db 329 VSSPFPN-----NRHP----- 340
QY 215 GTATVAGSTVNSGMWFGSGRDGLINASNGTIVQADHNDGAS--IVIOHTNGFVSSYI 272
Db 341 ----VTGRIKPKGVDSVSGTPTVAPADQTVKAVQAGARVYMLRRHREYQTVYM 396
QY 273 HKDAQVKTGPTVTRTGRIASMKQ--PSGAL-FFPRISNGYVDP 320
Db 397 HLSKSLVAGQTVKKGRIALSGNTGISTGPHLHFRHI--NGRAVNP 445

RESULT 39
S71558
probable cell wall-plasma membrane linker protein FRP precursor - rae
N/Alternate names: hybrid-proline-rich protein
C/Species: Brassica napus (rape)
C/Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C/Accession: S71558
R/Author: W.; Pallas, J.A.; Jenkins, G.I.
Plant Mol. Biol. 31, 771-781, 1996
A/Title: Transcripts of a gene encoding a putative cell wall-plasma membrane linker prote
in.
A/Reference number: S71558; MUID:96400032; PMID:8806408
A/Accession: S71558
A/Molecule type: DNA
A/Residues: 1-376 <GNO>
A/Cross-references: EMBL:X94976; NID:9155067; PIDN:CA64425.1; PID:9155068
A/Superfamily: hydroxyproline-rich glycoprotein
F/1-27/Domain: signal sequence #status predicted <STG>
F/28-376/Product: probable cell wall-plasma membrane linker protein FRP #status predicted
Query Match 7.8%; Score 131; DB 2; Length 376;
Best Local Similarity 42.3%; Pred. No. 0.041;
Matches 33; Conservative 1; Mismatches 32; Indels 12; Gaps 2;
QY 136 TPSPVAVQSRPVQOHPAVQKPTPPVYVKKPTPPVYVQGPAPVAPVTEAPVATG 193
Db 176 TPTPPVATPPTPTPTPTPPVATPPTPAPVATPPTPTPTPTPTPTPTPTPTPTPTPT 234
QY 194 SSGVWQFRYPVATNPV 211
Db 235 -----PPTPTPPV 243

RESULT 40
E75461
dehydrated cell wall protein - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C/Accession: E75461
R/Author: O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.D.; I
.; M.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036996; PMID:10567266
A/Accession: E75461
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-665 <WHI>
A/Cross-references: GB:AE001943; GB:AE00513; NID:96458624; PIDN:AAFI0484.1; PID:9645862
A/Experimental source: strain R1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:50:52 ; Search time 22 Seconds

(without alignments)
607.063 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670

Sequence: 1 MTVTAINSONQKPIKRLGL.....LFEFRISHNGVYDPLVTLK 322

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	298	17.8	297	1 NLPD_PSEAE	P45682 pseudomonas
2	238	14.3	251	1 YGER_ECOLI	Q46798 escherichia
3	219.5	12.1	405	1 LPPB_HAETN	P44833 haemophilus
4	210	12.6	373	1 NLPD_SALTI	O56131 salmonella
5	209	12.5	377	1 NLPD_SALDU	P39700 salmonella
6	209	12.5	377	1 NLPD_SALDU	P40827 salmonella
7	202.5	12.1	379	1 NLPD_SALTI	P33648 escherichia
8	194	11.6	337	1 LPPB_HAESO	P36685 haemophilus
9	146.5	8.8	439	1 ACWA_LACLA	O9c1t4 lactococcus
10	146.5	8.8	484	1 P60_LISMO	P21171 listeria iv
11	146	8.7	524	1 P60_LISIV	O01837 listeria iv
12	144.5	8.7	523	1 P60_LISSE	O01838 listeria be
13	140.5	8.4	511	1 P60_LISGR	O01835 listeria gr
14	139.5	8.4	467	1 P60_LISIN	O01836 listeria in
15	137.5	8.2	524	1 P60_LISWE	O01839 listeria we
16	135	8.1	410	1 YIRB_HABIN	P44864 haemophilus
17	132	7.9	475	1 YERB_HABIN	P44693 haemophilus
18	130.5	7.8	437	1 ACWA_LACLC	Q46803 lactococcus
19	126.5	7.6	374	1 PRUB_ACHLY	P27458 actinobact
20	124	7.4	1742	1 GUNA_CALSA	P22534 caldocellum
21	123.5	7.4	666	1 MUR2_ENTHR	P29346 enterococu
22	118	7.1	633	1 LAI4_YEAST	Q12446 saccharomy
23	118	7.1	1192	1 RTN4_HUMAN	O9nec3 homo sapien
24	116.5	7.0	282	1 HES1_MOUSE	P35428 mus musculu
25	116	6.9	334	1 NLPD_BUCAL	P57493 buchnera ap
26	116	6.9	346	1 PRP1_LYCES	Q00451 lycopersico
27	115	6.9	97	1 NLPD_YEREN	P47764 yerishia en
28	114.5	6.9	1	1 MOZ_HUMAN	O92794 homo sapien
29	114	6.8	281	1 HES1_RAT	O46666 rattus norv
30	113.5	6.8	1331	1 MANB_CALSA	P22333 caldocellum
31	109.5	6.6	280	1 HES1_HUMAN	Q14469 homo sapien
32	109.5	6.6	1181	1 MYSC_DICDI	P42522 dictyosteli
33	109	6.5	907	1 VGP3_EBV	P03200 Epstein-bar

34	108.5	6.5	283	1 EXTN_SORBI	P24152 sorghum bic
35	108.5	6.5	555	1 GPI_CHLRE	Q9FPG6 chlamydomon
36	108.5	6.5	3591	1 PRAB_BORBE	P12255 bordetella
37	108	6.5	1593	1 AT12_HUMAN	P58397 homo sapien
38	107.5	6.4	671	1 ALYS_ENTFA	P37710 enterococu
39	107.5	6.4	777	1 PRZE_MYXXA	P18769 myxococcu
40	107.5	6.4	2167	1 SHK1_RAT	Q9wv48 rattus norv
41	107	6.4	444	1 SLAP_LACAC	P35829 lactobacilli
42	107	6.4	553	1 ODP2_ALCEU	O59098 alcaligenes
43	106.5	6.4	3176	1 CA36_HUMAN	P12111 homo sapien
44	106	6.3	532	1 SPG7_DICDI	P22698 dictyosteli
45	106	6.3	672	1 HMW3_MYCPN	Q50360 mycoplasma

ALIGNMENTS

RESULT 1
ID NLPD_PSEAE STANDARD; PRT; 297 AA.
AC P45682;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipoprotein nlpd/lppb homolog precursor.
GN P45623.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Collier L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
-1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
-1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
-1- This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; D26134; BA05130.1; -;
CC EMBL; AE004782; AA07011.1; -;
CC MEROPS; M37. UPW; -;
CC InterPro; IPR002482; LysM.
CC InterPro; IPR002886; Peptidase_M37.
CC InterPro; IPR000437; Prok_lipoprot.
CC Pfam; PF01476; LysM; 1.
CC Pfam; PF01551; Peptidase_M37; 1.
CC SMART; SMO0257; LysM; 1.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; FALSE_NEG.

KW Inner membrane; Lipoprotein; Signal; Complete proteome.
 FT SIGNAL 1 22 PROBABLE
 FT CHAIN 23 297 LIPOPROTEIN NLPPB_HOMOLOG.
 FT LIPID 23 23 N-ACYL DIGLYCERIDE (PROBABLE).
 SQ SEQUENCE 297 AA; 30835 MW; 2CDDP26F4CBCT04 CRC64;

Query Match 17.8%; Score 298; DB 1; Length 297;
 Best Local Similarity 28.0%; Pred. No. 8.5e-14;
 Matches 91; Conservative 50; Mismatches 106; Indels 78; Gaps 12;

20 LIFGVITTCILAGCASKP-----TYNSTSGSGSHRTSGGGLAIGSOVITDSQGVNR 72
 25 LLLGAVVCSLLAACSSPPGKVKVDRNSAPAAARRTPVTSQ-----Q 68
 73 YQYKQDVTYSKIAQRYGLMWRIGHINNNSSYTYTGWMLTMSGDLKVRERSISSGVN 132
 69 YVRRGDTLYSIAFRFGWDMKALAANGIAPPTYIQVQ----- 107
 133 TAHPSPVAVQ--SSRPVQCHPAVOKPTPVV--VVKKPTPTPVVQ-----QPAVAP 184
 108 -----AIQFGGRASTQ--PSVAKNTPVAVAPVAKPTPVPAVSTSVPAKAP-APA 156
 185 VTEAPPATSGSGVQVQ-----FRYPVATNPVVRFRGTATVAGSTVTSNGMFGSGDGL 238
 157 STTPPSGATPVVAVGAVGAMVPAAGT--LIGRF-----ASNGSLNGIDIAQGLQGP 209
 239 INNSNGVTYQADHNDGAS--IVIOHTNGFVSYYHIKDAQVKTGDTVRTGORISMGN 296
 210 VLAASGTVVYAGSGLRGVELIITHNETVYVSAYGHNRLLVREGQVYKVGQSIAMFMS 269
 QY 297 QPSGALFERRISRGVYVDLTVL 321
 DB 270 TGTDRYKLFHFRKROKRPDLPQYL 294

RESULT 2

YGER_ECOLI STANDARD; PRT; 251 AA.
 ID YGER_ECOLI
 AC Q46798;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Hypoetical lipoprotein Yger precursor.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NC NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; Pubmed=9278503;
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1232-1244 (1997).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U28375; AAA83046.1; ALT INIT.
 DR EMBL; AEO00369; AAC75903.1; ALT INIT.
 DR EcoGene; EG13048; YGER.
 DR InterPro; IPR002483; LysM.

DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SMO0257; LysM; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Hypoetical protein; Membrane; Lipoprotein; Signal;
 KW Complete proteome.

FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 251 HYPOTHETICAL LIPOPROTEIN YGER.
 FT LIPID 26 26 N-ACYL DIGLYCERIDE (POTENTIAL).
 SQ SEQUENCE 251 AA; 26564 MW; FCD6235D9790475 CRC64;

Query Match 14.3%; Score 238; DB 1; Length 251;
 Best Local Similarity 24.9%; Pred. No. 9.4e-10;
 Matches 78; Conservative 46; Mismatches 103; Indels 86; Gaps 8;

16 KRLGIIFGVITTCILAGCASKPTYNSTSGSHRTSGGGLAIGSOVITDSQGVNRYQV 75
 9 KSLGIWMLISVGLLAGC-----SGSKSDPTGYSG-----SVYTV 44
 76 KQDVTYSKIAQRYGLMWRIGHINNNSSYTYTGWMLTMSGDLK---VRESISSGV 131
 45 KRGDLYRISRTTGSVELRLNGISPPYIEVGQKLT--GGAKSSITRKSTAKSTT 102
 132 NTAA--TPSPVAVQSSRPVQCHPAVOKPTPVVVKKPTPTPVVQOPAPVAPVTAPE 190
 103 KTAASVTPSSAVPKSSWPPVQRCWLMPTTGKVM-----PY 138
 191 ATGSSGVQVFRYPVATNPVVRFRGTATVAGSTVTSNGMFGSGDGLINNSNGVTYQ 250
 139 STADG-----NKGIDISARGTPIYVAGGKVVV 169
 QY 251 DHNDGAS--IVIOHTNGFVSYYHIKDAQVKTGDTVRTGORISMKNQPSGALFEFR 308
 DB 170 GNGLRGVNLMIKHSEDIITAYAHNDPMLVNNQSVKAGKXIAWGSTDAASVRLHQI 229
 QY 309 SRNGVYVDLTVL 321
 DB 230 RYRATADLPDLYL 242

RESULT 3

LPPB_HAEIN STANDARD; PRT; 405 AA.
 ID LPPB_HAEIN
 AC P44833;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Outer membrane antigenic lipoprotein B precursor.
 GN LPPB OR H10706.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 NC NCBI_Taxid=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; Pubmed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kierkegaard A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenny K., Sutton R., Liu L.-I., Giordano A., Kelley J.M.,
 RA Scott J.D., Shiley R., Liu L.-I., Giordano A., Kelley J.M.,
 RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Usterback T.R., Hanna M.C., Nguyen T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fitchman J.U., Funtmann J.L., Georgiades N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT rd.";
 RL Science 269:496-512 (1995).
 CC -1- FUNCTION: MAY BE A VIRULENCE DETERMINANT.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor.

CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32753; AAC2363.1; -
 DR MEROPS: M37.UPM; -
 DR TIGR: H10706; -
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR002886; Peptidase_M37.
 DR Pfam: PF01476; LysM; 1.
 DR Pfam: PF01551; Peptidase_M37; 1.
 DR SMART: SM00257; LysM; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR Outer membrane; Lipoprotein; Signal; Antigen; Virulence;
 KW Complete proteome.
 FT SIGNAL 1 17 PROBABLE.
 FT CHAIN 18 405 OUTER MEMBRANE ANTIGENIC LIPOPROTEIN B.
 FT LIPID 18 18 N-ACYL DIGLYCERIDE (PROBABLE).
 SQ SEQUENCE 405 AA; 42505 MW; 4134FB6B93EF57F CRC64;
 Query Match 13.1%; Score 219.5; DB 1; Length 405;
 Best Local Similarity 27.8%; Pred. No. 3e-08;
 Matches 76; Conservative 46; Mismatches 114; Indels 37; Gaps 10;
 Oy 71 NRYQYKQSDTVSKIAQRYGLMREIGHININSSYTIYQWLTMSGDLKVRERSISG 130
 Db 145 NRYQYKQSDTVSKIAQRYGLMREIGHININSSYTIYQWLTMSGDLKVRERSISG 202
 Oy 131 VNTAHTSPVAVQSRPPVQOHPANQKPTP-----PVVVKKPTPTPPV 175
 Db 203 -----VKQPAVTAATATPVK--PAVTV--TPGANGTQIGSDGITIGPKSEAGTSPSVEA 254
 Oy 176 QQPAPVAPVTEAPATGSSGVWQFRYPVATNPVVRBFGTA--TVAGSTV---SNGMW 230
 Db 255 TSSIVQTVSSVNNANSTPINSNVVA--PI-ASN-VVWQMPISGNIIGFSSITDGNKIID 309
 Oy 231 FSGRDGDLINMSNACTVIOADHNDGAS--IVIQHTNGFVSASYIHKDAQVKTGPTVETG 288
 Db 310 ISGSGQAVKAAARIVYAGNALGYNGLIITIKNDPFLSYANDKILVADQGEVAG 369
 Oy 289 QRASMKRQPSGALFERRISNGYVDPVTL 321
 Db 370 QDIAMGSSGTNTVLAHFEIRYKGSVDPVRYL 402
 RESULT 4
 NLPD_SALTI STANDARD; PRT; 373 AA.
 ID NLPD_SALTI AC Q56131;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lipoprotein nlpd precursor.
 GN NLPD OR STY3050.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 CX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churchill C., Mungall K.L., Bentley S.D., Holden M.R., Sebatina M.,
 RA Baker S., Baeham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,

RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE OF 255-373 FROM N.A.
 RC STRAIN=Ty2a;
 RX MEDLINE=95220644; PubMed=7705608;
 RA Robbe-Saule V., Coynault C., Norel F.;
 RT "The live oral typhoid vaccine Ty2a is a rpos mutant and is
 RT susceptible to various environmental stressors.";
 RL FEMS Microbiol. Lett. 126:171-176(1995).
 RN [3]
 RP SEQUENCE OF 269-373 FROM N.A.
 RC STRAIN=Ty2;
 RX MEDLINE=99118315; PubMed=9919662;
 RA Robbe-Saule V., Norel F.;
 RT "The rpos mutant allele of Salmonella typhi Ty2 is identical to that
 RT of the live typhoid vaccine Ty2a.";
 RL FEMS Microbiol. Lett. 170:141-143(1999).
 CC -1- FUNCTION: MAY BE INVOLVED IN STATONARY-PHASE SURVIVAL.
 CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
 CC anchor (Potential).
 CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL627276; CAD06031.1; -
 DR EMBL: X81641; CA57297.1; -
 DR EMBL: Y17610; CA476806.1; -
 DR MEROPS: M37.UPM; -
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR002886; Peptidase_M37.
 DR Pfam: PF01476; LysM; 1.
 DR Pfam: PF01551; Peptidase_M37; 2.
 DR SMART: SM00257; LysM; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Inner membrane; Lipoprotein; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 25
 FT CHAIN 26 373 LIPOPROTEIN NLPD.
 FT LIPID 26 26 N-ACYL DIGLYCERIDE (BY SIMILARITY).
 FT DOMAIN 199 246 4 X 7 AA APPROXIMATE REPEATS.
 FT REPEAT 199 205 1.
 FT REPEAT 221 227 2.
 FT REPEAT 233 239 3.
 FT REPEAT 240 246 4.
 SQ SEQUENCE 373 AA; 39183 MW; DD075BB440098FF7 CRC64;
 Query Match 12.6%; Score 210; DB 1; Length 373;
 Best Local Similarity 23.2%; Pred. No. 1.2e-07;
 Matches 84; Conservative 56; Mismatches 138; Indels 84; Gaps 13;
 Oy 30 LAGCASK-----PYNSTSGSGSHRTSG-----SGGL-----AIGS 60
 Db 23 LAGTSSSNPPAPVTSVDSGSSSTNGMLTPPKXGATPQQAPOIQPVPTQGMQT 82
 Oy 61 QVITDSQ-----GVF-----NRYQYKQSDTVSKIAQRYGLMREIGHI 98
 Db 83 QPVTQSPQVQWENGRIVNRYQYGNIPKGSYTGSGTYVTKKGTLLFYIAMITGNDPRDAQR 142
 Oy 99 NNNLSYTIYQGL-----TWSGDLKVRERSISGNTAHTPSPAVAVQSRPPV- 149
 Db 143 NSIAPYSLNVGQLTVQGNASGPTPTGNAITQDAAQGGVTVRSAGNSTVAVASOPTTT 202
 Oy 150 -----QOHPAVQKPTP--PVVVVKKPTPTPPVQOHPAPVAPVTEAPATGSSGVWQFR 201

```

DB      203 YSESGEGSANKMKNPKPAGTVAPVTAFTVSTTEPNASTSTSAF-----ISAMR 255
QY      202 YPVGATNPVVRFRGTATVAGSTVTSNGMWFSGRGGDLINASAGTVIAQDHNDGCA--I 259
DB      256 WP--TDGKVIENFG-ASEGG-----NKGIDIASGKGAIVATADGRVYVAGNALRGVNL 308
QY      260 VIQHTNGFVSSYTHIKDAQVKTGDPVTRGQRIASMKQPSGAALFEPRISRNQVYVDELT 319
DB      309 IINKNDYLSAYAHNDTMLVREOQEVKAGKIAITMGSTGSTRLHFIRYKGSVNLPR 368
QY      320 VL 321
DB      369 YL 370

RESULT 5
NLPD_SALDU STANDARD; PRT; 377 AA.
AC P39700;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipoprotein nlpD precursor.
GN NLPD.
OS Salmonella dublin.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxId=98360;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RC STRAIN=lane;
RX MEDLINE=9938748; PubMed=973354;
RA Paesold G., Krause M.;
RT "Analysis of rpos mRNA in Salmonella dublin: identification of
RT multiple transcripts with growth-phase dependent variation in
RT transcript stability."
RT J. Bacteriol. 181:1264-1268(1999).
RN [2]
RP SEQUENCE OF 284-377 FROM N.A.
RC STRAIN=lane;
RA Krause M.W., El-Gedaily A.;
RT Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (Potential).
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ006131; CA06881.1; -
DR EMBL: X82129; CA57639.1; -
DR MEROPS: M37.1UP; -
DR InterPro: IPR002483; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR InterPro: IPR000437; Prok_LipProt.
DR Pfam: PF01476; LysM; 1.
DR SMART: SM00257; LysM; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Inner membrane; Lipoprotein; Repeat; Signal.
FT CHAIN 1
FT SIGNAL 25
FT LIPID 26 377 BY SIMILARITY.
FT DOMAIN 26 26 LIPOPROTEIN NLPD.
FT REPEAT 203 250 N-ACYL DIGLYCERIDE (BY SIMILARITY).
FT REPEAT 203 209 4 X 7 AA APPROXIMATE REPEATS.
FT REPEAT 225 231 2.

```

```

FT REPEAT 237 243 3.
FT REPEAT 244 250 4.
SQ SEQUENCE 377 AA; 39671 MW; DC7F59B943F86512 CRC64;
Query Match 12.5%; Score 209; DB 1; Length 377;
Best Local Similarity 23.0%; Pred. No. 1.5e-07;
Matches 84; Conservative 56; Mismatches 138; Indels 88; Gaps 13;

QY 30 LAGGASK-----PYNSTSSGSHRTSG-----SGGLA----- 57
DB 23 LAGTSSSNPPAPVTSVDSGSSSVTNSGMLITPPKMGATQTPQAPQIPVQRPVQ 82
QY 58 -IGSOVITDQ-----GVP-----NRVOKGDDTVSKTAORYGLMRE 94
DB 83 PMQTPVTEQPVCMENGRIVYNNROYNI PKGSYTGSTIYVKKGDTLFYIAMTGNDFR 142
QY 95 IGHINNLANSYTYTGQWL-----TMSGLKXREBSISSGVNFAHPPSPVAVQSR 146
DB 143 LAQRNLSAPYSLNAGTLOVGNASGMPITGNAITQADAAQGVTRSAQNSTVANAASQ 202
QY 147 PPV-----QHPAVQKPT--PPVVVYKPTPPPVVQOPAPVAPVTEAPFATGSSGV 197
DB 203 PTIYSESGEGSANKMKNPKPAGTVAPVTAFTVSTTEPNASTSTSAF-----I 255
QY 198 MQFRYPVGTATNPVVRFRGTATVAGSTVTSNGMWFSGRGGDLINASAGTVIAQDHNDGCA 257
DB 256 SAMRWP--TDGKVIENFG-ASEGG-----NKGIDIASGKGAIVATADGRVYVAGNALRGV 308
QY 258 S--IVQHTNGFVSSYTHIKDAQVKTGDPVTRGQRIASMKQPSGAALFEPRISRNQVY 315
DB 309 GNLIIKNDYLSAYAHNDTMLVREOQEVKAGKIAITMGSTGSTRLHFIRYKGSV 368
QY 316 DPLFTVL 321
DB 369 NPLRVL 374

RESULT 6
NLPD_SALTY STANDARD; PRT; 377 AA.
AC P40827;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipoprotein nlpD precursor.
GN NLPD OR STM2925.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxId=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=2153448; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RT Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 280-377 FROM N.A.
RC STRAIN=ATCC 14028;
RX MEDLINE=9436857; PubMed=8086465;
RA Prince R.W., Fang F.C., Libby S.U.;
RT Cloning and sequencing of the gene encoding the rpos (katF) sigma
RT factor from Salmonella typhimurium 14028.
RT J. Bacteriol. Biophys. Acta 1219:196-200(1994).
RN [3]
RP SEQUENCE OF 298-377 FROM N.A.
RC STRAIN=C52;

```

RA	MEDLINE=95050252; PubMed=7961444; Kowarz L., Cymnalt C., Robbe-Saulle V., Norcl F.;
RT	"The Salmonella typhimurium katF (tpos) gene: cloning, nucleotide sequence, and regulation of gpvR and spvABCD virulence plasmid genes.";
RL	J. Bacteriol. 176:6852-6860(1994).
CC	-1- FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.
CC	-1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid anchor (Potential).
CC	-1- SIMILARITY: BELONGS TO THE E. COLI NUPD / HAEMOPHILUS LPPB FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AE008833; AAL21805.1; --.
DR	EMBL; U05011; --; NOT_ANNOTATED_CDS.
DR	EMBL; X77752; --; NOT_ANNOTATED_CDS.
DR	MEROPS; M37.UPW; --.
DR	StyGene; SG10513; nlpD.
DR	InterPro; IPR002886; Peptidase_M37.
DR	InterPro; IPR000437; Prot_lipoproc.
DR	Pfam; PF01551; Peptidase_M37; 1.
DR	PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM	Inner membrane; Lipoprotein; Repeat; Signal; Complete proteome.
FT	SIGNAL
FT	CHAIN
FT	LIPID
FT	DOMAIN
FT	REPEAT
FT	REPEAT
FT	REPEAT
FT	REPEAT
FT	CONFPEAT
FT	SEQUENCE

```

Query Match      12.5%; Score 209; DB 1; Length 377;
Best Local Similarity 23.0%; Pred. No. 1.5e-07;
Matches 84; Conservative 56; Mismatches 138; Indels 88; Gaps 13.

QY      30 LAGCASK-----PTYNSTSGSGSHRTSG-----SCGLA-----57
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      23 LAGCTSSNPAPATSVSDSSSSNNSSGMLTPPEKMGATTGQTRQAPOIQVQRPVYQ 82
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      58 -TSGOVTDSQ-----GVP-----NRVYKQGTVSKIAQRYGLNRE 94
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      83 PMQOPVPEQOVOMENGRIVYNNROYGNPKPSYTGSGSYTYVKEKGTPLVYIAMITDNDRD 142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      95 IGHINNLSSTYIYQWL-----TLMGDLKVRERSSISGVNTAHTSPVAVQSSR 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      143 LAQRNRSIAPSLANGQTLLOYGNASGRTITGNAATQADAAQGVYVTSAGQNSTYAVASQ 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      147 PPV-----QQHPAQKPT--PPVYVYKPTPPVYVQAPAPAPVYTEAPFATGSSGV 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      203 PLITYEGSGRQSANMKLPNNKKPACTVVTAPYAPTAFTVSTPEBNASTSTAP-----I 255
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      198 MQFRPVPVATPPVVRFGTATVAGSTVSNCGMWFSGRGDDLMAINAGTVIOADHNMGCA 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      256 SAMRPP--TDCKVIENFQ--ASEGQ---NKGIIDLAGSKQALVAVADRVRVYVYAGALNGY 308
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      258 S--IYIQTNGFVSSYIHIKDAQVYTGDTGTVTRGRIASMKRQPSGALFEFRISNGYV 315
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      309 GNLIIITKNDYLSAYANDMTLVAREQGEVAKGAKIATMGSTGTSTLHFIPIRYKGSV 368
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      316 DPLVTVL 321
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      369 NPLRYL 374
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7			
NLPD_ECOLI	STANDARD;	PRT;	379 AA.
ID NLPD_ECOLI			
AC P33648;			
DT 01-FEB-1994 (Rel. 28, Created)			
DT 01-FEB-1994 (Rel. 28, Last sequence update)			
DT 16-OCT-2001 (Rel. 40, Last annotation update)			
DE Lipoprotein nlpd precursor.			
GN NLPD OR B2742.			
OS Escherichia coli.			
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC Escherichia.			
OX NCBI_TaxID=562;			
<div> <div>[1]</div> <div>SEQUENCE FROM N.A.</div> </div>			
RP STRAIN=MP180;			
RC MEDLINE=94179096; PubMed=8132457;			
RA Ichikawa J.K., Li C., Fu J.C., Clarke S.,			
RT "A gene at 55 minutes on the Escherichia coli chromosome encodes a			
RL lipoprotein with unusual amino acid repeat sequences."			
J. Bacteriol. 176:1630-1638(1994).			
[2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=K12 / MG1655;			
RX MEDLINE=97426617; PubMed=9278503;			
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.T.,			
RA Mau B., Shao Y.;			
RT "the complete genome sequence of Escherichia coli K-12.";			
RL Science 277:1453-1474(1997).			
[3]			
RP SEQUENCE OF 99-379 FROM N.A.			
RC STRAIN=K12 / DH1;			
RX MEDLINE=94268497; PubMed=8208244;			
RA Takayanagi Y., Tanaka K., Takahashi H.;			
RT "Structure of the 5' upstream region and the regulation of the rpos			
RL gene of Escherichia coli."			
Mol. Gen. Genet. 243:525-531(1994).			
CC -1- FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.			
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid			
CC anchor (Potential).			
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.			
CC -----			
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration -			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).			
CC CC			
DR EMBL, L07869; AAA17875.1; -;			
DR EMBL, U29579; AAA6252.1; -;			
DR EMBL, AE000358; AAC5784.1; -;			
DR EMBL, D17549; BAA04487.1; -;			
DR MEROPS; M37.0PM; -;			
DR SWISS-2DPAGE; P33648; COLI.			
DR Ecocore; EG12111; nlpd.			
DR InterPro; IPR002482; LysM.			
DR InterPro; IPR002886; Peptidase_M37.			
DR Pfam; PF01476; LysM_1.			
DR Pfam; PF01551; Peptidase_M37; 1.			
DR SMART; SMO0257; LysM; 1.			
DR PROSITE; PS00013; PROKR_LIPOPROTEIN; 1.			
KM Inner membrane; Lipoprotein; Repeat; Signal; Complete proteome.			
FT SIGNAL	1	25	
FT CHAIN	26	379	LIPROTEIN NLPD.
FT LIPID	26	26	N-ACYL DIGLYCERIDE.
FT DOMAIN	66	97	4 X 8 AA TANDEM REPEATS OF Q-Q-P-Q-I-Q-P-V.
FT REPEAT	66	73	1-1.
FT REPEAT	74	81	1-2 (APPROXIMATE).


```

CC -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006264; AAK04370.1; -
DR InterPro: IPR002901; Amidase_4.
DR InterPro: IPR000423; Flag_P193.
DR Pfam: PF01476; LysM; 3.
DR Pfam: PF01832; Amidase_4; 1.
DR PRINTS: PRO1002; FLAGL50.
DR SMART: SM00047; LysM; 1.
DR SMART: SM00257; LysM; 3.
DR SMART: SM00257; LysM; 3.
DR Hydrolase: Glycosidase; Bacteriolytic enzyme; Cell wall;
DR Cell division; Septation; Signal; Repeat; Complete proteome.
DR SIGNAL 1 57
DR CHAIN 1 57
DR REPEAT 243 285
DR REPEAT 323 365
DR REPEAT 397 439
DR DOMAIN 287 291
DR DOMAIN 370 374
DR SEQUENCE 439 AA; 46592 MW; AE4551B4D04CB499 CRC64;

Query Match 8.8%; Score 146.5; DB 1; Length 439;
Best Local Similarity 23.6%; Pred. No. 0.0034;
Matches 69; Conservative 39; Mismatches 107; Indels 77; Gaps 10;

Oy 26 TTCILAGCASKPTYNST-----SGSGSHRTSGSGGLAIGSOVITDSCGVNPR 72
Db 184 TAAAGKATATDPNVCASLNRRIISQYNLTRFDGASAGTSNGSG-STATVNNNSNTSSTT 242
Oy 73 YOVKQGDVYSKIAQRYGKLNMRIGHNINLNSVTYTGQWLTWLGDKLVKRSISGVN 132
Db 243 YTVSGDILWGLISQKIGISVAGIOSANNLKRST-VYIQGLVLTSSSSSNTSSTSGN 301
Oy 133 TAAHTPSPAVAVOSSRPVQOHPAVOKRTPPVVKKRTPTPPVVQOPAVAPVPT-----EA 188
Db 302 SAGITTPPT-----SVTPAKPASQTTIKVKSGLTLMGLSV 336
Oy 189 PPAVSSGAVMORRYVGAATNPPVRRFGATVAGSTVTTSNGKMFSGRDDLINASVAGTVI 248
Db 337 KYKTTIAQLKSMNHLNSDTIPFGQMLIVSOSAGSSSSSTG-----SSSAST-- 382
Oy 249 QADHMDGASIVIGHNGFVSSYHIIKQAOVKTGPTV-----RTGRIASMK 295
Db 383 SSTSNSAAS-----NTSIH-----KVKKGDLTLMGLSOKSGSPIASIK 420

RESULT 10
P60_LISMO STANDARD; PRT; 484 AA.
ID P60_LISMO
AC P21171; Q03493;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein p60 precursor (Invasion-associated protein).
GN IAP OR LMO0582.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-49.
RC STRAIN=EGD / Serovar 1/2a;
RX MEDLINE=90256283; PubMed=2111287;
RA Koehler S., Leimweister-Waechter M., Chakraborty T., Ioteppeich F.,
RA Goebel W.;

```

```

RT "The gene coding for protein p60 of Listeria monocytogenes and its
RT use as a specific probe for Listeria monocytogenes.";
RL Infect. Immun. 58:1943-1950 (1990).
RN [2]
RP SEQUENCE FROM N.A., AND DISCUSSION OF SEQUENCE.
RC STRAIN=Mackness / Serovar 1/2a;
RX MEDLINE=93094153; PubMed=1459966;
RA Hubert A., Kuhn M., Goebel W., Koehler S.;
RT "Structural and functional properties of the p60 proteins from
RT different Listeria species.";
RL J. Bacteriol. 174:8166-8171 (1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangoul L., Buchrieser C., Rusniok C., Anend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Darivar A., Deloux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kuypkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueder T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852 (2001).
CC -1- FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
CC INVASION OF NONPROFESSIONAL PHAGOCYTIC CELLS BY LISTERIA.
CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
CC BINDING.
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLP / LISTERIA P60 FAMILY.
CC -1- SIMILARITY: CONTAINS 2 LYSM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52268; CA36509.1; -
DR EMBL: M80351; AAA25280.1; -
DR EMBL: AL581975; CAC98661.1; ALT_INIT.
DR PIR: A41487; A41487.
DR L1a11a1c; LMO00582; -
DR MEROPS; C40.0PM; -
DR InterPro: IPR002482; LYSM.
DR InterPro: IPR000644; SH3_P60.
DR InterPro: IPR003646; SH3_Fac.
DR Pfam: PF00877; NLP_P60; 2.
DR Pfam: PF01476; LysM; 2.
DR SMART: SM00257; LysM; 2.
DR SMART: SM00287; SH3; 1.
KW Repeat; Signal; Complete proteome.
FT SIGNAL 1 27
FT CHAIN 1 27
FT DOMAIN 311 355
FT REPEAT 30 72
FT REPEAT 203 245
FT VARIANT 94 94
FT VARIANT 167 167
FT VARIANT 196 196
FT VARIANT 326 331
FT SEQUENCE 484 AA; 50587 MW; 3CC0F90591E14E0F CRC64;

Query Match 8.8%; Score 146.5; DB 1; Length 484;
Best Local Similarity 20.6%; Pred. No. 0.0038;
Matches 81; Conservative 46; Mismatches 121; Indels 145; Gaps 14;

Oy 25 ITTCILAGCASKPTYNSTSGSGSHRTSGSGGLA--IGSOVITD----- 65

```



```

DB 102 IITSLKGG--TKVVTETTESNMHKTITNDGKTGVNGKYLTDKAVSTPVPATQVEKKT 159
OY 66 --SGVP-----NRQYQKQDPTVSKIAQRYG 89
DB 160 TTQQAAPAAETKEVKQTQATTPAPKAEKTEPVDONATTHAVKSGDITWALSXKIG 219
OY 90 LMRREIGHINNLSSTYITYGQMLTWSGDLKVERRSISSGVNTAHTSPVAVQSGRPV 149
DB 220 VSVQDIMSNNLSSS-SIVVGQKLA-----KQNTATAPKEVKT----- 259
OY 150 QQHPAVQKTPPVV-----VKKETPTPPVQGPAPVAPVTEAFPATGSSGVNQ 199
DB 260 -EAPAAEKQAAEVVKNENTNTATTEKKTATQ--QQTAPKAPTEAAKAPAPAFSTNTVA 315
OY 200 FRYPVGA-----TNPVRRFGTATVAGSTVTSNGMFPSGRDGLINASNAQTVIQ 249
DB 316 NKTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 375
OY 250 ADHND-----GASIVQHTNGF-VSSYHIKDAQVKTGPT 284
DB 376 AQKHIDKAVSWGNGPTTFDCGYTKVPAKAGISLPTSGAQVASTRTISSQAKPGDL 435
OY 285 VRTGQRIAMKQKPSGALFEF--RISRGVVY 315
DB 436 V-----PFDYQSGISHYGIIV 451

RESULT 11
P60_LISTIV STANDARD; PRT; 524 AA.
ID P60_LISTIV
AC 001837;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein p60 precursor (Invasion-associated protein).
GN IAP.
OS Listeria ivanovi.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_Taxid=1638;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384582; Pubmed=1514809;
RA Buber A., Koehler S., Goebel W.;
RT "The homologous and heterologous regions within the iap gene allow
RT genus- and species-specific identification of Listeria spp. by
RT polymerase chain reaction."
RT Appl. Environ. Microbiol. 58:2625-2632(1992).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=93094153; Pubmed=1459966;
RA Buber A., Kuhn M., Goebel W., Koehler S.;
RT "Structural and functional properties of the p60 proteins from
RT different Listeria species."
RT J. Bacteriol. 174:8166-8171(1992).
RN [3]
RP FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
RN INVASION OF NONPROFESSIONAL PHAGOCYTTIC CELLS BY LISTERIA
RN -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOLYCAN
RN BINDING.
RN -1- SIMILARITY: BELONGS TO THE E. COLI NLPC / LISTERIA P60 FAMILY.
RN -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
RN [4]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation-
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [5]
RP EMBL; M80350; AAA25284.1;
RN MEROPS; C40.0PM;
DR Interpro; IPR002482; LYSM.

```

```

DR Interpro; IPR000064; NLPC_P60.
DR Interpro; IPR003646; SH3_Dac.
DR Pfam; PF00877; NLPC_P60; 1.
DR Pfam; PF01476; LYSM_3.
DR SMART; SM00287; LYSM; 3.
DR SMART; SM00287; SH3b; 1.
KM Repeat; Signal.
FT SIGNAL.
FT CHAIN.
FT REPEAT.
FT REPEAT.
FT REPEAT.
FT REPEAT.
SQ SEQUENCE 524 AA; 54159 MW; E22D579429817144 CRC64;

Query Match 8.7%; Score 146; DB 1; Length 524;
Beat Local Similarity 24.6%; Pred. No. 0.0045;
Matches 78; Conservative 40; Mismatches 95; Indels 104; Gaps 17;

OY 25 IITCLAGCAKPYNTSTSGSHRTS-GSG-----GLAIGQV-----ITD 65
DB 100 IITSLKGG--TKVVTETTESNMHKTITNDGKTGVNGKYLTDKAVSTPVPATQVEKKT 157
OY 66 --SGVP-----NRQYQKQDPTVSKIAQRYG 92
DB 158 AQAAPAAETKEVKQSTPATVLPRAETKETPAPVDTASTVTSVSGDITWALSXKIGTSV 217
OY 93 REIGHINNLSSTYITYGQMLTWSGDLKVERRSISSGVNTAHTSPVAVQSGRPVQ 151
DB 218 QNIMSNNLSSS-SIVVGQVLA-----VKQEA-----AKTAPFAEVKTEAPFAEK 262
OY 152 H---PAVKETPPVAVVKKPTPTPVV-VQGPAPVAPVTEAFPATGSSGVNQFRYPVQAT 207
DB 263 QVSAFVAKENTQTTAKKEVFPQKQNTQAPAAQAKP-APAPAFVNTNASSYTVKSGDT 321
OY 208 NPVARRFGTATVAGSTVTS-----NGMFPSGRDGLINASN-----AGTVIQADHND 254
DB 322 LSKT-----ATFTGVTSKIKALNG-----LNSDNLQVQVLYKVKGTVPANTNS 366
OY 255 D-GASIVQHTNGFPVS 270
DB 367 NSNATAPPTNTNTNTSN 383

RESULT 12
P60_LISTSE STANDARD; PRT; 523 AA.
ID P60_LISTSE
AC 001838;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein p60 precursor (Invasion-associated protein).
GN IAP.
OS Listeria seeligeri.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_Taxid=1640;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384582; Pubmed=1514809;
RA Buber A., Koehler S., Goebel W.;
RT "The homologous and heterologous regions within the iap gene allow
RT genus- and species-specific identification of Listeria spp. by
RT polymerase chain reaction."
RT Appl. Environ. Microbiol. 58:2625-2632(1992).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=93094153; Pubmed=1459966;
RA Buber A., Kuhn M., Goebel W., Koehler S.;
RT "Structural and functional properties of the p60 proteins from
RT different Listeria species."
RT J. Bacteriol. 174:8166-8171(1992).
RN [3]
RP FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
RN INVASION OF NONPROFESSIONAL PHAGOCYTTIC CELLS BY LISTERIA.
RN -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOLYCAN

```



```

CC BINDING.
CC -1- SIMILARITY: BELONGS TO THE E.COLI NLPC / LISTERIA P60 FAMILY.
CC -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80353; AAA25286.1; -.
CC DR MEROPS; C40.UF0; -.
CC DR InterPro; IPR002482; LYSM.
CC DR InterPro; IPR000064; NLPC_P60.
CC DR InterPro; IPR003646; SH3_Fac.
CC DR Pfam; PF00877; NLPC_P60; 1.
CC DR Pfam; PF01476; LYSM; 3.
CC DR SMART; SM00257; LYSM; 3.
CC DR SMART; SM00287; SH3b; 1.
CC KW Repeat; Signal.
CC FT SIGNAL 1 27 BY SIMILARITY.
CC FT CHAIN 28 523 PROTEIN P60.
CC FT REPEAT 30 72 LYSM 1.
CC FT REPEAT 200 242 LYSM 2.
CC FT REPEAT 320 362 LYSM 3.
CC SQ SEQUENCE 523 AA; 53844 MW; 89341210D20DF6B4 CRC64;

Query Match 8.7%; Score 144.5; DB 1; Length 523;
Best Local Similarity 20.4%; Pred. No. 0.0057;
Matches 93; Conservative 58; Mismatches 131; Indels 173; Gaps 21;

QY 1 MNTVTAISNQKPKIKRLG-----IFGVTTTCLLGCAPRYNYSGGSGSR-TSG 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 69 LQVTEVASEKTEKYSATWLNVRSGAGVDSIVTSLKQ3--TKVIVESTEANGMKITYG 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 53 SG-----GLAIGSOVITDSQGV----- 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 127 ESKTGYVANGKYLGNVNTSPATPRVKEQETTOAPAOQTEVKAQTPATTEKDAVET 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 -----NRVYKQGDIVSKIAQRYGLNREIGHINNINSSYTYITGQMLTMSGD 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 187 KTTAPAVDNTATTTHTVSGDITLWLSVKYGAVIDLMSNNLSSS-SIVYG-----N 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 LKVRERSISSGVNT-----AHTPSVAVOSSRPVQOHPAVKPPPVVVKKPTPT--- 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 239 IAVKQ--SAAKNTAPKAEAKTEAPAEKQTAAPVVK--STNTSTTTTVKKEETTEKQ 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 172 -----PVVQOPAPVAPV--TEAPFATGSSG----- 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 293 TSTTAPAPQAAKAPAPAPVNTNASTYVKSGLDLGKIASFTGTVSKIKALNGLTSDN 352
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 -----VMOFRYPVAGATNPVRRFGTAIVAGSTVTSNGMWFSGRDDLINASVAGT--- 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 353 LQVGVDLKVKGVAPATNT-----NTAT-APTNTNTNN-----TSSSTSTSPSKNN 396
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 -VIQADHMDGASITYQHT-----NGRVSSYTIKKAQV---KTG 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 397 NTNQSSSSSSASATIAEQKILGRAYSGANGPTTFPCSGP-TSYVFAQSGITLPRISG 455
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 DTVRTGORTIASMKNOPSGALPEF--RISRNGVY 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 456 AQYASTTKVSESEAPGDLVFPDYSGIAHVGIV 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DE Protein p60 precursor (Invasion-associated protein).
GN IAP.
OS Listeria grayi (Listeria murrayi).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_Taxid=1641;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92384582; PubMed=1514809;
RA Hubert A.; Koehler S.; Goebel W.;
RT "The homologous and heterologous regions within the iap gene allow
RT genus- and species-specific identification of Listeria spp. by
RT polymerase chain reaction.";
RL Appl. Environ. Microbiol. 58:2625-2632(1992).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=93094153; PubMed=1459966;
RA Hubert A.; Kuhn M.; Goebel W.; Koehler S.;
RT "Structural and functional properties of the p60 proteins from
RT different Listeria species.";
RL J. Bacteriol. 174:8166-8171(1992).
CC -1- FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
CC INVASION OF NONPROFESSIONAL PHAGOCYTTIC CELLS BY LISTERIA.
CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
CC BINDING.
CC -1- SIMILARITY: BELONGS TO THE E.COLI NLPC / LISTERIA P60 FAMILY.
CC -1- SIMILARITY: CONTAINS 2 LYSM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80352; AAA25285.1; -.
CC DR MEROPS; C40.UF0; -.
CC DR InterPro; IPR002482; LYSM.
CC DR InterPro; IPR000064; NLPC_P60.
CC DR InterPro; IPR003646; SH3_Fac.
CC DR Pfam; PF00877; NLPC_P60; 1.
CC DR Pfam; PF01476; LYSM; 3.
CC DR SMART; SM00257; LYSM; 3.
CC DR SMART; SM00287; SH3b; 1.
CC KW Repeat; Signal.
CC FT SIGNAL 1 27 BY SIMILARITY.
CC FT CHAIN 28 511 PROTEIN P60.
CC FT REPEAT 177 219 LYSM 1.
CC FT REPEAT 299 342 LYSM 2.
CC SQ SEQUENCE 511 AA; 53181 MW; 2FD0ED174E3810D0 CRC64;

Query Match 8.4%; Score 140.5; DB 1; Length 511;
Best Local Similarity 26.6%; Pred. No. 0.011;
Matches 47; Conservative 24; Mismatches 83; Indels 23; Gaps 5;

QY 73 YQVKGDPVSKTARQYGLNREIGHINNINSSYTYITGQMLTMSGDKYVERSISSGVN 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 177 YKVSGLDITLWLSVKYGVQVQKLTLENNLSSS-SIVYG-----TLVVEAQAAPPT 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 TAHTPSVAVOSSRPVQOHPAVOKPTPPVV-VVKKPTPPPVVQOPAPVAPVTEAPFA 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 229 TVKQAAPKAVAPKQEVQOTAPAKQKAPAKETVYKRAVSKPATPTAKPAVEQKAS 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 TGSSGVQOFRYPVAGATNPVRRFGTAIVAGSTVTSNGMWFSGRDDLINASNACTVI 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 289 T-----PAVDTAATYKVN-----GDSLGIKIASLFFVSVDLNNMNNLNATI 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
P60_LISGR STANDARD; PRT; 511 AA.
ID P60_LISGR
AC 001835;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

RESULT 14
P60_LISIN STANDARD; PRT; 467 AA.
ID P60_LISIN
AC 001836;
DT 01-APR-1993 (Rel. 25, Created)

DT 15-JUN-2002 (Rel. 41, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE protein p60 precursor (Invasion-associated protein).
 GI IAP OR L1N0591.
 OS *Listeria innocua*.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; *Listeria*.
 NCBI_TaxID=1642;
 RX SEQUENCE FROM N.A.
 RX STRAIN=serovar 6b;
 RX MEDLINE=92384582; PubMed=1514809;
 RA Rubert A., Koehler S., Goebel W.,
 "The homologous and heterologous regions within the iap gene allow
 RT genus- and species-specific identification of *Listeria* spp. by
 RT polymerase chain reaction.";
 RL Appl. Environ. Microbiol. 58:2625-2632(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 1122 / Serovar 6a;
 RC MEDLINE=21537279; PubMed=11679669;
 RA Glaeser F., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
 Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 Gautier L.-M., Keerst U., Kreft J., Kuhn M., Kunst F., Kuxaraki G.,
 Jones L., Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
 Nardone G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 Remmel B., Rose M., Schlueter T., Simoes N., Tietz A.,
 Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.,
 "Comparative genomics of *Listeria* species.";
 RL Science 294:849-852(2001).
 RN [3]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=93094153; PubMed=1459966;
 RA Rubert A., Kuhn M., Goebel W., Koehler S.,
 "Structural and functional properties of the p60 proteins from
 RT different *Listeria* species.";
 RL J. Bacteriol. 174:8166-8171(1992).
 CC -1- FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
 CC -1- INVASION OF NONPROFESSIONAL PHAGOCYTIC CELLS BY *LISTERIA*.
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
 CC -1- BINDING.
 CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPC / *LISTERIA* P60 FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 LYSM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M80349; AAA25283.1; -;
 CC EMBL: AL596165; CAC95823.1; ALT_INIT.
 CC List: L1N0591; -;
 CC DR MEROPS: C40.0pw; -;
 CC DR InterPro: IPR002482; LYSM.
 CC DR InterPro: IPR000064; NLPC_P60.
 CC DR InterPro: IPR003646; SH3_Bac.
 CC DR Pfam: PF00877; NLPC_P60; 2.
 CC DR Pfam: PF01476; LYSM; 2.
 CC DR SMART: SM00257; LYSM; 2.
 CC DR SMART: SM00287; SH3; 1.
 CC DR Repeat: Signal; Complete proteome.
 CC FT SIGNAL 1 27
 CC FT CHAIN 28 467
 CC FT DOMAIN 330 343
 CC FT REPEAT 30 72
 CC FT REPEAT 201 243
 CC FT VARIANT 240 240
 CC L -> P (IN SEROVAR 6B).

FT VARIANT 328 328 K -> KNTNTNTNTNTNT (IN SEROVAR 6B).
 SQ SEQUENCE 467 AA; 49044 MW; 0F40B031BD2175C1 CRC64;
 Query Match 8.4%; Score 139.5; DB 1; Length 467;
 Best Local Similarity 21.1%; Pred. No. 0.011;
 Matches 75; Conservative 51; Mismatches 143; Indels 87; Gaps 14;
 QY 25 ITTCLACASKEPTNSTSGSGSHRTSGSGGLA--IGSOVITD-----65
 DB 101 ILTSLKGG--TKVYETTESNGMKHTYNDGKTGVNKKYLDKATSTPPVKKQVKETT 158
 QY 66 -----SOGVP-----NRYOQSDTVSKIAORYGL 90
 DB 159 QQVKAETAKTEVKOPTQOTAPAKAEKTEAPVVDONATHNVKSGDITMALSVKYG 218
 QY 91 NKREIGHINNLSSTITTYGMLTMSGDLKVRERSISSGVNTATTPPVAQSRPVO 150
 DB 219 SVQDITMSNNLSSS-SIVGQKLA---IKOPTVAPKAEK-KIQAPAEKQIAPAVK 271
 QY 151 QHPAVQKPTPPVAVVKKPTPPVVOQAPVAPVTEAPFATSGSGVQFRYPVQATNPV 210
 DB 272 ENSNNTAT---TEKKETATE---QOTTAPQAPAPAPAPSTNTKNTNTNTNTAS 324
 QY 211 VRRFGTAVAGSTVTSNGWFSGRDDILNSNAGTVIQAD-----HNMDSASIVIGT 264
 DB 325 TPKSKNT 379
 QY 265 NGFVSSYHIKDA--QVKTGTVRTGORIASMKQPSGALFEF--RISRGVYV 315
 DB 380 SGF-TKVFYAKSGISLPRTSGAQVASTRISBSQAKPDVLFYDYGSGISHVIGIV 434
 RESULT 15
 ID P60 LISWE STANDARD; PRT; 524 AA.
 AC 001839;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein p60 precursor (Invasion-associated protein).
 GN IAP.
 OS *Listeria welshimeri*.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; *Listeria*.
 NCBI_TaxID=1643;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92384582; PubMed=1514809;
 RA Rubert A., Koehler S., Goebel W.,
 "The homologous and heterologous regions within the iap gene allow
 RT genus- and species-specific identification of *Listeria* spp. by
 RT polymerase chain reaction.";
 RL Appl. Environ. Microbiol. 58:2625-2632(1992).
 RN [2]
 RP SEQUENCE FROM N.A., AND DISCUSSION OF SEQUENCE.
 RX MEDLINE=93094153; PubMed=1459966;
 RA Rubert A., Kuhn M., Goebel W., Koehler S.,
 "Structural and functional properties of the p60 proteins from
 RT different *Listeria* species.";
 RL J. Bacteriol. 174:8166-8171(1992).
 CC -1- FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
 CC -1- INVASION OF NONPROFESSIONAL PHAGOCYTIC CELLS BY *LISTERIA*.
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
 CC -1- BINDING.
 CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPC / *LISTERIA* P60 FAMILY.
 CC -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; M80354; AAA5281.1; ALT_SEQ.
DR EMBL; M80348; AAA5281.1; ALT_SEQ.
DR HSSP; P23931; 180G.
DR MEROPS; C40 UPW; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR000064; NLP_C_P60.
DR InterPro; IPR003646; SH3_Bac.
DR Pfam; PF00877; NLP_C_P60_1.
DR Pfam; PF01476; LysM_3.
DR SMART; SM00257; LysM_3.
DR SMART; SM00287; SH3b_1.
DR Repeat; Signal.
KM SIGNAL 1 27 BY SIMILARITY.
FT CHAIN 28 524 PROTEIN P60.
FT REPEAT 30 72 LYSM 1.
FT REPEAT 198 240 LYSM 2.
FT REPEAT 316 356 LYSM 3.
SQ SEQUENCE 524 AA; 54086 MW; 228AF5DC4A254C9 CRC64;

```

```

Query Match 8.2%; Score 137.5; DB 1; Length 524;
Best Local Similarity 22.9%; Pred. No. 0.017;
Matches 74; Conservative 37; Mismatches 103; Indels 109; Gaps 14;

```

```

QY 73 YQVKGQDTVSKIAQRYGLNWRERIGHINNINSYTYITGQWLTMSGDKVERSSISGVN 132
DB 198 HTVKSQDTIMALSVKYGASVDLMGMNLS--SIYVQ-----KIAVKS 242
QY 133 TAHFSPVAVQSSRPVQOHPVQKPT--PPVY-----PVKK-----PTPPPVQ 176
DB 243 AAKTAPAPAEVKT-----EAPAVEKETSTPVKKENTNTTVKKEVTQTQTNTTKAPQAA 297
QY 177 QAPVAPV-----TEAPFATSSGVMQFRYPVGAATNPVRREGTAT----- 218
DB 298 KPAPAPAPVNTNTASTYTVKSG-----DSISKIANTGTISVKIKALNLTSDNL 349
QY 219 -----VAGSTVTSNGMFGSGRDDLINASNAGT-----VIQADHMDGA 257
DB 350 QVGTVLAKYGVPTNTNNSNTTAPTNTNNSNTSNTSTPSKNTNTNNGSSNSASA 409
QY 258 SIVIOHT-----NGFVSYIHIDA--QVKTGPTVATGGRISM 294
DB 410 SALIAEAKHUKAYSWGNGPTTPDCSGF--TSYVFAKSGISLPTSGAQVASTRISBS 468
QY 295 KNPSGALFEF--RISRGVYV 315
DB 469 QAKEDLVFPDYGSGIAHVGYV 491

```

RESULT 16

```

YIBP_HAEIN
ID YIBP_HAEIN STANDARD; PRT; 410 AA.
AC P44864;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0756.
GN HI0756.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,

```

```

RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- SIMILARITY: STRONG, TO E.COLI YIBP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.ibt-sib.ch/announce/
CC or send an email to license@ibt-sib.ch).
CC -----
DR EMBL; U32760; AAC22415.1; -.
DR MEROPS; M37.UPW; -.
DR TIGR; HI0756; -.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37_1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 410 AA; 45983 MW; B28D93E4E813BCA9 CRC64;

```

```

Query Match 8.1%; Score 135; DB 1; Length 410;
Best Local Similarity 25.0%; Pred. No. 0.02;
Matches 63; Conservative 36; Mismatches 111; Indels 42; Gaps 10;

```

```

QY 74 QVKGQDTVSKIAQRYGLNWRERIGHINNINSYTYITGQWLTMSGDKVRRRSISGVN 133
DB 188 QKKQQAALQKQOQEHQ-----STTNEIKNLALDQDQNT-----LKANEQALRQETQR 236
QY 134 AHFSPVAVQSSRPV--QOHPVQKPTPPVVKKPTTPPVVQOAPVAPVTEAPFAT 192
DB 237 AEQAAREGEKERRALAQQAEEKRT-----SKRYQFTVGERQLN 278
QY 193 GSGGV---WQFRYPVGAATNPVRRPFGTAVAGSTVTSNGMFGSGRDDLINASNAGTVI 248
DB 279 STSGIGAAKQKQSLPVSGS--ILHTFG--SIQAGEVMKGVAVIGASAGTPVKALIAAGFVI 334
QY 249 QADH--NMDGASIVIOHTNGFVSYIHIDKADQVKTGPTVATGGRISMKN--QPSGALFE 305
DB 335 LAGTNGYGVYIVVHGETDLSLGFNOVASKVQGLVSAGQVIAQVNGTGEISRSALY 393
QY 306 FRISRGVYVDP 317
DB 394 FGISRKGPVNP 405

```

RESULT 17

```

YEBP_HAEIN
ID YEBP_HAEIN STANDARD; PRT; 475 AA.
AC P44653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0409.
GN HI0409.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,

```

RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RL Science 269:496-512 (1995).
 CC -1- FUNCTION: COULD BE INVOLVED IN CELL WALL DEGRADATION OR
 CC FORMATION (BY SIMILARITY).
 CC -1- SIMILARITY: STRONG, TO E. COLI Y8A.
 CC -1- SIMILARITY: TO STAPHYLOCOCCUS LYSTOABIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, U32724; AAC22068.1; -.
 CC DR MEMOPS; M37; UPW; -.
 CC DR TIGR; H10409; -.
 CC DR InterPro; IPR002886; Peptidase_M37.
 CC DR Pfam; PF01551; Peptidase_M37; 1.
 CC KW Hypochemical protein; Cell wall; Complete proteome.
 CC SEQUENCE 475 AA; 53255 MW; 143C10F92233939D CRC64;
 Query Match 7.8%; Score 132; DB 1; Length 475;
 Best Local Similarity 23.0%; Pred. No. 0.037;
 Matches 67; Conservative 37; Mismatches 83; Indels 104; Gaps 12;
 QY 72 RYGVKGD--TVSKIAORYGLNMRERIGHNNL-----NSSYTI-----YTQG 111
 Db 217 RKEVLAGEIGNLSNLSREKGLDTRQISQSLNMQVSLKLLKKGQFALIVSRYYLD 276
 QY 112 WTLNMGDLKYRERRSSGVNTAHTSPVAVOS-----SRPVQCHRA 154
 Db 277 KLT---GQGNVELARISSGGKNTY-----AVQANGRYVQOGETLKGARFPLQROAR 328
 QY 155 VQKPTPPVYVVKPTPPVYVQOPAPVAPVYTAFTATSSGVNORRYPVGATNPVYRRR 214
 Db 329 VSSPFND-----NRHP----- 340
 QY 215 GATVAGSTVTSNGMFGSGDGLINASNAVTIQAIDHNDGAS--IVIOHTNGFVSSYI 272
 Db 341 ---VTGRIRPHKGVDFVSQGFPIAPADGTVKAVYQAGAGRYVMLHGREYQTVYM 396
 QY 273 HIRDAQVKTGDIYRTQORIASMNQ--PSGAL--FEFRISRNGVYDPLTV 320
 Db 397 HLRKSLVKAQOTVKKERIALSGNTGISTGPHLHYEPHI--NGRAVAVPLTV 445
 RESULT 18
 ACMA LACTIC STANDARD; PRT; 437 AA.
 AC Q48603; O52362;
 DT 16-OCT-2001 (Ref. 40, Created)
 DT 16-OCT-2001 (Ref. 40, Last sequence update)
 DT 16-OCT-2001 (Ref. 40, Last annotation update)
 DE Probable N-acetylmuramidase precursor (EC 3.2.1.17) (Peptidoglycan
 DE hydrolase) (Autolysin) (lysosyme).
 GN ACMA.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_taxid=1359;
 RN RP
 RN SEQUENCE FROM N.A.
 RX STRAIN=M61363;
 RX MEDLINE=95189733; PubMed=7883712;
 RA Buiet G., Kok J., Leenhouts K.J., Dabrowska M., Venema G.,
 RA Haandrikman A.J.;
 RT "Molecular cloning and nucleotide sequence of the gene encoding the
 RT major peptidoglycan hydrolase of Lactococcus lactis, a muramidase
 RT needed for cell separation."
 RT J. Bacteriol. 177:1554-1563 (1995).

RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=2250;
 RC MEDLINE=21015606; PubMed=11131071;
 RA Govindasamy-Lacey S., Gopal P.K., Sullivan P.A., Pillidge C.J.;
 RA "Varying influence of the autolysin, N-acetyl muramidase, and the cell
 RT envelope proteinase on the rate of autolysis of six commercial
 RT Lactococcus lactis cheese starter bacteria grown in milk."
 RT J. Dairy Res. 67:585-596 (2000).
 CC -1- FUNCTION: HYDROLYSES THE CELL WALL OF L. LACTIS AND
 CC M. LISOCEITRICUS. REQUIRED FOR CELL SEPARATION DURING GROWTH.
 CC CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
 CC heteropolymers of the prokaryotic cell walls.
 CC CC -1- DOMAIN: THE LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN
 CC PEPTIDOGLYCAN BINDING.
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, U17696; AAC3367.1; -.
 CC DR EMBL; AF036720; AAB93629.1; -.
 CC DR InterPro; IPR002901; Amidase_4.
 CC DR Pfam; PF01476; LysM; 3.
 CC DR Pfam; PF01832; Amidase_4; 1.
 CC DR SMART; SM00047; LysM; 3.
 CC KW SMART; SM00257; LysM; 3.
 CC KW Hydrolase, Glycosidase, Bacteriolytic enzyme; Cell wall;
 CC Cell division; Septation; Signal; Repeat.
 CC STGNL 1 57
 FT CHAIN 1
 FT REPEAT 245 287
 FT REPEAT 321 363
 FT REPEAT 395 437
 FT REPEAT 27 27
 FT CONFLICT 31 31
 FT CONFLICT 64 64
 FT CONFLICT 155 155
 FT CONFLICT 190 190
 FT CONFLICT 231 231
 FT CONFLICT 285 285
 FT CONFLICT 316 316
 FT CONFLICT 320 320
 FT CONFLICT 368 368
 FT CONFLICT A -> T (IN REF. 2).
 SQ SEQUENCE 437 AA; 46564 MW; 5C905633BD5D28B CRC64;
 Query Match 7.8%; Score 130.5; DB 1; Length 437;
 Best Local Similarity 24.1%; Pred. No. 0.043;
 Matches 69; Conservative 39; Mismatches 119; Indels 59; Gaps 12;
 QY 26 TCIILAGASKRTVST-----SGSGHRTSGGGLAIGSOVITDSCGVN- 71
 Db 184 TIALGRVAYDPSYASLARIISOYNLRFQDASSAGNTNSG--STTITINNNSGNS 240
 QY 72 ---RYGVKGDIVSKIAORYGLNMRERIGHNNLSYTYTGQWTLNMGDLKYRERIS 128
 Db 241 SSTTVVSGDITLWGISQRYGISAQIOSANLAKST-IITGQKLVLGSG--ASTNS 295
 QY 129 GSVNTAHTSPVAVOSRPPVQOPAPVOKPTPPVYVVKPPT--PVYQOPAPVAPPV 186
 Db 296 GGSNNNS-----ASTPTTTSVTPA--KPTSQITVKKSGDITWALSVKKTSTIAQKS 345
 QY 187 EAPFATGSGGWQ---FRYPVATNPVVRFGTATVAGSTVTSNGMFGSGDGLINASN 243
 Db 346 WNLHLSDDITVIGQNLIVSQAASNP-----STGSGSTATNNNSNSTSNSNMAIHKV 398

QY 244 AGTVIOADHNDGASIVIOHTNGFVSYIHKDAQVKTGPTVQ 289
DB 399 KQDTIMGLSOKSGSPI-----ASIKAMNHL-----SSDTLLIQ 432

RESULT 19
PRLB_ACHLY STANDARD; PRT; 374 AA.
AC P27458;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Beta-lytic metalloendopeptidase precursor (EC 3.4.24.32) (Beta-lytic protease).
OS Achromobacter lyticus.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Achromobacter.
OX NCBI_TaxID=224;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 196-220.
RC STRAIN=M497-1.
RX MEDLINE=91035265; PubMed=2228973;
RA Li S.L., Noriocka S., Sakiyama F.;
RT "Molecular cloning and nucleotide sequence of the beta-lytic protease gene from Achromobacter lyticus."
RL J. Bacteriol. 172:6506-6511 (1990).
CC -1- CATALYTIC ACTIVITY: Cleavage of N-acetyl[muramyl]-Ala, and of the inulin B chain at 23-Gly-|-Phe-24 > 18-Val-|-Cys(SO3)H).
CC -1- CONDUCTOR: BINDS ONE ZINC ION.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M23.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M60896; AAA21906.1; -.
CC PIR; A37151; LYXYL; -.
CC MEROPS; M23.001; -.
CC InterPro; IPR000841; Blytic_Mendpep.
CC PRINTS; PR00933; BLYTICPTASE.
CC Hydrolase; Metalloprotease; Zinc; Zymogen; Signal.
CC SIGNAL 25 195
CC PROPEP 1 24
CC CHAIN 196 374 BETA-LYTIC METALLOENDOPEPTIDASE.
CC DISULFID 261 307 BY SIMILARITY.
CC DISULFID 351 364 BY SIMILARITY.
CC METYL 316 316 ZINC (POTENTIAL).
CC METYL 318 318 ZINC (POTENTIAL).
CC SEQUENCE 374 AA; 40084 MW; 431E51B84575DE14 CRC64;

Query Match 7.6%; Score 126.5; DB 1; Length 374;
Best Local Similarity 25.9%; Pred. NO. 0.066; Indels 49; Gaps 7;
Matches 50; Conservative 29; Mismatches 65;
QY 117 SGLDKVRRSRISSGVNTAHTSPVAVOSRPVQOHAPVOKTTPVVVKKPTPPVVO 176
DB 154 AGGRAARRRVPAGL-----RPVQRTAGCGGFGP-----LR 186
QY 177 GPAPVAPVTEAPPTGSSGVQFPYVGAIVPVRRRTATVAGS-----TVTSNG 228
DB 187 QGRP-----GRAVVSNGILQFPFPGASMTV---GGAHNTSGVYPMSSLSMRKG 236
QY 229 WPFSGRDDLINASAGTVIOADHNDGASIVIOHTNGFVSYIHKDAQVKTGPTVQ 288
DB 237 GKGSGNQNGVWVSASASGFKR--HSCRAEIV--HTGWSITYYHLMNIQYNTGANVSMN 292
QY 289 ORIASMKQPSGA 301

DB 293 TAINPANTQAOA 305
RESULT 20
GUNA_CALSA STANDARD; PRT; 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A) (Cellulase A).
GN CEBA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'O V.S., Saul D.J., Bergquist P.L.;
RT "Cela, another gene coding for a multidomain cellulase from the extreme thermophile Caldocellum saccharolyticum."
RL Appl. Microbiol. Biotechnol. 43:291-296 (1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuchli E., Usamat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'."
RL Appl. Environ. Microbiol. 57:694-700 (1991).
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE CELLULOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.
CC -1- PRT: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; L32742; AAA91086.1; -.
CC EMBL; M36063; AAA72860.1; -.
CC PIR; A43745; A43745.
CC HSSP; P26221; ITP4.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR001701; GH_9.
CC InterPro; IPR000556; Glyco_hydro_48.
CC Pfam; PF00759; Glyco_hydro_9; 1.
CC Pfam; PF00942; CBM_3; 3.
CC Pfam; PF02011; Glyco_hydro_48; 1.
CC PRINTS; PR00844; GLYDRLASE48.
CC ProDom; PD011903; Glyco_hydro_48; 1.
CC ProDom; PD011903; Glyco_hydro_48; 1.
CC PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
CC SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 1742 ENDOGLUCANASE A.

```

FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1142 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F069A2123EED07 CRC64;

Query Match
Best Local Similarity 23.5%; Pred. No. 0.56;
Matches 76; Conservative 51; Mismatches 130; Indels 66; Gaps 17;

7 INSGQKPKIKGLI-----FGVITTCILAGCASKPTVNSTSGSGSHRTSG-----SGGL 56
DB VNNQGMFARATNKLKFRFVDSLIRAGVSPNQLTSTVYNQAGKVSFPVWDSNNI 561
QY 57 AIGSOVITDSQGV---PNRYVQGDVTSKIAQRYGLNWRIGHINNIN-----S 103
DB 562 ---YIIVDFGTGLYFGGQDKYKKEVQFRIAPQNVQW-----DNSNDYFQDIKVS 612
QY 104 SVTITGOMLTMSGDLV-REKSSISGVN---TAHPSPVAVQSSPPVQGHAVQKP 158
DB 613 SSSVTKKTIPIYXDDIKVWGEPSGVSPTPTASVPTPTPTPTPTPTPTPTPTPT 671
QY 159 TPVVVVKKPP---TPPVQCPAPVAPVTEAPFAGSSGVMGFRVAVGATNVARF 214
DB 672 TPTVATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 726
QY 215 -----GATVAGSVTNSGAMFGSGRQDLINASNAQVITQADHNMDCASITVQHTNGFV 268
DB 727 LKVNSSGSSIDLSRYLTR-YWYV-----VDEGRQSAI-SDMAQIGASNV---TFKTV 775
QY 269 --SSYIHKDAQVKTQDITVITQO 289
DB 776 KLSSSVSGADYVLEIGFKSGAQO 798

RESULT 21
MUR2_ENTR STANDARD; PRT; 666 AA.
AC P39046;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Muramidase-2 precursor (EC 3.2.1.17) (1,4-beta-N-acetylmuramoylhydrolase) (Peptidoglycan hydrolase) (Pg-hydrolase-2) (Lysozyme).
OS Enterococcus hirae.
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID:1354;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-73.
RC STRAIN=ATCC 9790;
RX MEDLINE=92165737; PubMed=1347040;
RA Chu C.-P., Kariyama R., Daneo-Moore L., Shockman G.D.;
RT "Cloning and sequence analysis of the muramidase-2 gene from
RT Enterococcus hirae.";
RL J. Bacteriol. 174:1619-1625 (1992).
RN [2]
RP FUNCTION.
RC STRAIN=ATCC 9790;
RX MEDLINE=89327152; PubMed=2753858;
RA Dolinger D.L., Daneo-Moore L., Shockman G.D.;
RT "The second peptidoglycan hydrolase of Streptococcus faecium ATCC 9790
RT covalently binds penicillin.";
RL J. Bacteriol. 171:4355-4361 (1989).
RN [3]
RP FUNCTION.

```

```

RC STRAIN=ATCC 9790;
RA Del Mar Lleo M., Canepari P., Satta G.;
RT "Thermosensitive cell growth mutants of Enterococcus hirae that
RT elongate at non-permissive temperature are stimulated to divide by
RT parental autolytic enzymes.";
RL J. Gen. Microbiol. 139:13099-13117 (1993).
CC -1- HYDROLYTIC ACTION OF MURAMIDASE-1, WHICH REQUIRES BINDING OF THE
CC ENZYME TO NONREDUCING ENDS OF GLYCAN CHAINS. HYDROLYSIS IN THE
CC MIDST OF GLYCAN CHAINS WOULD INCREASE THE NUMBER OF BINDING SITES
CC FOR MURAMIDASE-1. MAY FUNCTION IN FACILITATING SEPTUM FORMATION
CC AND CELL SEPARATION. ACTIVE ON M. LUTERUS CELL WALLS AND ON E. HIRAE
CC CELL WALL FRACTIONS, BUT NOT ACTIVE ON E. HIRAE INTACT CELL WALLS.
CC CAN COVALENTLY BIND PENICILLIN OF THE 1,4-BETA-LINKAGES BETWEEN N-
CC CATALYTIC ACTIVITY. HYDROLYSIS OF THE 1,4-BETA-LINKAGES BETWEEN N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan.
CC heteropolymers of the prokaryotes cell walls.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
CC BINDING.
CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77639; AAA24776.1; -.
DR PIR: A42296;
DR HSRP: P22629; 1SMD.
DR InterPro: IPR002901; Amidase_4.
DR InterPro: IPR002482; LysM.
DR Pfam: PF01476; LysM; 6.
DR SMART: SM00847; LysM; 6.
DR SMART: SM00257; LysM; 6.
DR SMART: SM00257; LysM; 6.
DR Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
DR Cell division; Septation; Repeat; signal.
KM SIGNAL
FT CHAIN 1 49
FT REPEAT 50 666 MURAMIDASE-2.
FT REPEAT 257 299 LYSM 1.
FT REPEAT 338 380 LYSM 2.
FT REPEAT 414 456 LYSM 3.
FT REPEAT 489 531 LYSM 4.
FT REPEAT 565 607 LYSM 5.
FT REPEAT 623 665 LYSM 6.
SQ SEQUENCE 666 AA; 70670 MW; FFOAVFAFCD810BA3 CRC64;

Query Match
Best Local Similarity 31.6%; Pred. No. 0.21;
Matches 36; Conservative 21; Mismatches 40; Indels 17; Gaps 4;

QY 31 AGCAKPKPTVNSTG---SGSHRTSGSGGLAIGSVITDSQGVPNRYVQKQDITVSKIAQR 87
DB 304 AGSSTTNGNNAASGNTGNTSGTQATGA-----KYTKSGDSYWKIND 352
QY 68 YGLNMRREIGINNNSYTIYQGLVLTMSGDLKVRERSISSGVTAHTPSPVA 141
DB 353 HOISNQLIEMNINKNF-VYFGQGLVVKSSSA--SGSTNSTGTGNTSSNTA 403

RESULT 22
LAL7_YEAST STANDARD; PRT; 633 AA.
ID LAL7_YEAST
AC Q12446;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Proline-rich protein LAL7.

```

GN LAS17 OR YOR181W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Toth E. A.;
 RN Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RA Hughes B., Pohl T.M.;
 RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO S.POMBE WSP1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; D78487; BAA1386.1; -;
 DR SGP; 50005707; LAS17.
 DR InterPro; IPR000697; RanBpl_WASP.
 DR InterPro; IPR001960; WH1.
 DR Pfam; PF00568; WH1.1.
 DR Pfam; PF02205; WH2.1.
 DR SMART; SM00461; WH1.1.
 DR SMART; SM00246; WH2.1.
 FT DOMAIN 185 190 POLY-PRO.
 FT DOMAIN 323 329 POLY-PRO.
 FT DOMAIN 342 348 POLY-PRO.
 FT DOMAIN 352 358 POLY-PRO.
 FT DOMAIN 385 391 POLY-PRO.
 FT DOMAIN 427 431 POLY-PRO.
 FT DOMAIN 470 474 POLY-PRO.
 FT DOMAIN 503 507 POLY-PRO.
 FT DOMAIN 520 526 POLY-PRO.
 SQ SEQUENCE 633 AA; 67571 MW; 4488355563AA2645 CRC64;
 Query Match 7 1%; Score 118; DB 1; Length 633;
 Best Local Similarity 23.0%; Pred. No. 0.47;
 Matches 65; Conservative 34; Mismatches 104; Indels 80; Gaps 12;
 QY 37 PTYNSTSGSG-----SHRTSGSGGLAIGSQVITDSOG----- 68
 DB 329 PMRTTBSSGVRLPAPPPRRGRAPPPRRHRTVNTLNAGNSLIPQATGRGRAP 388
 QY 69 -VPNRQYKQDPTVSKIAQRYGLNREIGHINLNS-----YTYTGQMLTMSG 118
 DB 389 PPRASRPTPVNTMQONPQYVNSNRPFQYQNSNNSPPPPVTTFTLFPQ-WTAATG 447
 QY 119 DLKV-----KRSISSGVNTAHNTSPYAVGSSRPVQOHAVQCPPTPVVVKETPTP 172
 DB 448 QPAVLPONTQAPSOATVPPVAPPPPPASLPQSGQIPQAPSPAPIDPPLPSTTSAAPPPP 507
 QY 173 P-VVQOP---APVAPVTEAPFATSSGVNQFRYPVQATNVVRRFGATVAGSTVSN 227
 DB 508 AFLTQPPSGAPAPPPPPQNP-ATSTSG-----GGSFATTT 543
 QY 228 GMMFSGRDGDLIN---ASNAGTVIQADHN-MDGASIVTQHTNG 266
 DB 544 G--DAGRDALLASIRGAGGIGALRKVDKSQLDKPVLQLQEARG 584

AC Q9Y5U6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Reticulon 4 (Neurite outgrowth inhibitor) (Nogo protein) (Foccen)
 DE (Neuroendocrine-specific protein) (NSP) (Neuroendocrine specific
 DE protein C homolog) (RTN-x) (Reticulon 5) (My043 protein).
 GN RTN4 OR NOGO OR ASY OR KIA0886.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RX MEDLINE=20129242; PubMed=10667780;
 RA Prinija R., Moore S.E., Vinson M., Blake S., Morrow R., Christie G.,
 RA Michalovich D., Simmons D.L., Walsh F.S.;
 RT "Inhibitor of neurite outgrowth in humans.";
 RL Nature 403:383-384(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Brain;
 RX MEDLINE=21010696; PubMed=11126360;
 RA Tagami S., Eguchi Y., Kinoshita M., Takeda M., Tsujimoto Y.;
 RT "A novel protein, RTN-XS, interacts with both Bcl-XL and Bcl-2 on
 RT endoplasmic reticulum and reduces their anti-apoptotic activity.";
 RL Oncogene 19:5736-5746(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RX MEDLINE=20237542; PubMed=20773680;
 RA Yang J., Yu L., Bi A.D., Zhao S.-Y.;
 RT "Assignment of the human reticulon 4 gene (RTN4) to chromosome
 RT 2p14-->2p13 by radiation hybrid mapping.";
 RL Cytogenet. Cell Genet. 88:101-102(2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RA Jin W.-L., Ju G.;
 RT "Developmentally-regulated alternative splicing in a novel Nogo-A.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC TISSUE=Placenta, and Skeletal muscle;
 RA Ito T., Schwartz S.M.;
 RT "Cloning of a member of the reticulon gene family in human.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Fibroblast;
 RA Yutendo M.;
 RT "Isolation of a cell death-inducing gene.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Pluticary;
 RA Song B., Peng Y., Zhou J., Huang Q., Dai M., Mao Y.M., Yu Y., Xu X.,
 RA Luo B., Hu R., Chen J.;
 RT "Human neuroendocrine-specific protein C (NSP) homolog gene.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
 RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
 RA Yu J., Han L.H.;
 RT "Novel human cDNA clone with function of inhibiting cancer cell
 RT growth.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XII.

RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RT DNA Res. 5:355-364 (1998).
 RN (110)
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC TISSUE=Brain, Pancreas, Placenta, and Skeletal muscle;
 RA Strussberg R.;
 RN Submitted (NOV-2000) to the EMBL/Genbank/DBD databases.
 RN (111)
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RN MEDLINE=20499367; PubMed=11042152;
 RA Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,
 RA Sheng Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
 RA Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.,
 RA "Cloning and functional analysis of cDNAs with open reading frames for
 RT 300 previously undefined genes expressed in CD34+ hematopoietic
 RT stem/progenitor cells."
 RT Genome Res. 10:1546-1560(2000).
 RN (112)
 RP SEQUENCE OF 482-1192 FROM N.A. (ISOFORM 1/4).
 RC TISSUE=Brain;
 RA Mao Y.M., Xie Y., Zheng Z.H.;
 RN Submitted (MAY-1998) to the EMBL/Genbank/DBD databases.
 RN (113)
 RP SEQUENCE OF 186-1192 FROM N.A. (ISOFORM 1).
 RC TISSUE=Testis;
 RA Sha J.H., Zhou Z.M., Li J.M.;
 RN Submitted (JAN-2001) to the EMBL/Genbank/DBD databases.
 RN (114)
 RP TISSUE=Brain;
 RN MEDLINE=20129259; PubMed=10667797;
 RA Grandpre T., Nakamura F., Vartanian T., Strittmatter S.M.;
 RT "Identification of the Nogo inhibitor of axon regeneration as a
 RT Reticulum protein."
 RT Nature 403:439-444(2000).
 CC -1- FUNCTION: Potent neurite outgrowth inhibitor which may also help
 CC block the regeneration of the nervous central system in adults.
 CC Isoform 2 reduces the anti-apoptotic activity of Bcl-Xl and Bcl-2.
 CC This is likely consecutive to their change in subcellular
 CC location, from the mitochondria to the endoplasmic reticulum,
 CC after binding and sequestration.
 CC -1- SUBUNIT: Interacts with Bcl-Xl and Bcl-2.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum. Anchored to the membrane of the endoplasmic reticulum
 CC through 2 putative transmembrane domains.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1/RTN 4A/Nogo-A/RTN-XL (shown
 CC here), 2/RTN 4B/Nogo-B/RTN-XS/Foocen-M, 3/RTN 4C/Nogo-C/Foocen-
 CC S and 4; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Isoform 1 is specifically expressed in brain
 CC and testis and weakly in heart and skeletal muscle. Isoform 2 is
 CC widely expressed excepted for the liver. Isoform 3 is expressed in
 CC brain skeletal muscle and adipocytes. Isoform 4 is testis-
 CC specific.
 CC -1- SIMILARITY: CONTAINS 1 RETICULON DOMAIN.
 CC -1- CAUTION: Ref.11 sequence differs from that shown due to
 CC frameshifts in positions 1149 and 1156.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, AJ251383; CAB99248.1; -
 CC EMBL, AJ251384; CAB99249.1; -
 CC EMBL, AJ251385; CAB99250.1; -
 CC EMBL, AB040462; BAB18927.1; -
 CC EMBL, AB040463; BAB18928.1; -
 CC EMBL, AF148537; AAG12176.1; -
 CC EMBL, AF148538; AAG12177.1; -

DR EMBL, AF087901; AAG12205.1; -
 DR EMBL, AF320999; AAG40878.1; -
 DR EMBL, AF132047; AAD11021.1; -
 DR EMBL, AF132048; AAD11022.1; -
 DR EMBL, AB015639; BAB3712.1; -
 DR EMBL, AF077050; AAD27783.1; -
 DR EMBL, AF177332; AAG17976.1; -
 DR EMBL, AB020683; BAB74909.1; -
 DR EMBL, BC001035; AAH01035.1; -
 DR EMBL, BC007109; AAH07109.1; -
 DR EMBL, BC014366; AAH14366.1; -
 DR EMBL, BC012619; AAH12619.1; -
 DR EMBL, BC010737; AAH10737.1; -
 DR EMBL, AF125103; AAD39920.1; ALT_FRAME.
 DR EMBL, AF063601; AAG43160.1; ALT_INIT.
 DR EMBL, AF333336; AAK20831.1; ALT_INIT.
 DR Genew, HGNC:14085; RTN4.
 DR MIM, 604475; -
 DR InterPro; IPR003388; Reticulon.
 DR Pfam; PF02453; Reticulon; 6.
 DR PROSITE; PS50845; RETICULON; 1.
 KW Endoplasmic reticulum; Alternating
 FT DOMAIN 1 1018
 FT TRANSMEM 1019 1039
 FT DOMAIN 1040 1133
 FT TRANSMEM 1134 1154
 FT DOMAIN 1155 1192
 FT DOMAIN 1005 1192
 FT DOMAIN 30 47
 FT DOMAIN 143 148
 FT VARSPLIC 1 993
 FT VARSPLIC 994 1004
 FT VARSPLIC 58 289
 FT VARSPLIC 186 1004
 FT CONFLICT 107 107
 FT CONFLICT 135 135
 FT CONFLICT 438 458
 FT CONFLICT 564 564
 FT CONFLICT 684 699
 SQ SEQUENCE 1192 AA; 129930 MW; CDE239EBF31589CA CRG64;
 Query Match 7.1%; Score 118; DB 1; Length 1192;
 Best Local Similarity 39.6%; Pred. No. 0.96; Mismatches 33; Indels 14; Gaps 6;
 Matches 36; Conservative 8;
 QY 132 NTAAHPSPV---AVQSRPPVQGHFAVQKTPPPVVYK--KPTFPVVOQPAFVADPV 185
 DB 115 STVPAPSPPLAAVAVSPSKLPEDDEPPAPPPPPASVSPDAEPVWTPA---PAPAAPPS 171
 QY 186 T-EAPFATGSGVQWQ---FRYPVGATNPVVR 212
 DB 172 TPAAPKRRSGSGSVDELFLAP-AASEPVIR 201
 RESULT 24
 HES1_MOUSE
 ID HES1_MOUSE STANDARD; PRT; 282 AA.
 AC P35428;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DS Transcription factor HES-1 (Hairy and enhancer of split 1).
 GN HES1 OR HES-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=94148977; PubMed=7906273;
 RX Takebayashi K., Sasai Y., Sakai Y., Watanabe T., Nakanishi S.,

CC Asteridae; euasteride I; Solanales; Solanaceae; Solanum.
 CC NCBI_TaxID=4081;
 RN [1]_TaxID=4081;
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv VPMT Cherry; Tissue=fruit;
 RX MEDLINE=92119262; PubMed=1731999;
 RA Sales Y., Machin R., Kengsbuch D., Grunseem W., Barg R.;
 RT "DNA sequence of the tomato fruit expressed proline-rich protein gene
 TRP-F1 reveals an intron within the 3 untranslated transcript";
 RL Plant Mol. Biol. 18:407-409(1992).
 RN [2]
 RP SEQUENCE OF 34-346 FROM N.A.
 RC STRAIN=cv Arava;
 RX MEDLINE=91329722; PubMed=1868217;
 RA Sales Y., Machin R., Grunseem W., Barg R.;
 RT "Sequence coding for a novel proline-rich protein preferentially
 expressed in young tomato fruit";
 RL Plant Mol. Biol. 17:149-150(1991).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X61395; CAA43666.1; -;
 DR EMBL: X57076; CAA40361.1; -;
 DR HSSP: P24337; 1HYP
 DR InterPro: IPR001768; Try/amy1_inhbr.
 DR Pfam: PF00224; TRYF_alpha_amy1_1
 SQ SEQUENCE 346 AA; 36375 MW; 60458452FEF16E0 CRC64;
 Query Match Best Local Similarity 6.9%; Score 116; DB 1; Length 346;
 Matches 33; Conservative 6; Mismatches 32; Indels 12; Gaps 5;
 QY 136 TPSPVAVQSSRPVQGHAPVQKPTPPV--VVKKP-TTPPVVQGPAPVAPVTAPPA 191
 DB 158 TTPPPPVV--HPVTPRP--PSRPVSPPIVPTTPPVVSPPIITPPPIVSPPV 212
 QY 192 TGSSGVQGFY---PVGATNPVV 211
 DB 213 PNPVVVIPPVVPSPVTPPIV 235
 RESULT 27
 NLPD YEREN STANDARD; PRT; 97 AA.
 AC P47764;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Lipoprotein nlpd (fragment).
 GN NLPD
 OS Yersinia enterocolitica.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia
 NCBI_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WI024 / Serotype O:9;
 RX MEDLINE=95247270; PubMed=7728993;
 RA Irlate M., Steiner I., Cornells G.R.;
 RT "The rps gene from Yersinia enterocolitica and its influence on
 expression of virulence factors";
 RL Infect. Immun. 63:1840-1847(1995).
 CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
 anchor (potential).
 CC -1- SIMILARITY: BELONGS TO THE E.COLI NLPD / HAEMOPHILUS LPPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U16152; AAC43390.1; -;
 DR MEROPS: M37 UPW; -;
 DR InterPro: IPR002886; Peptidase M37.
 DR InterPro: IPR000437; Prok_lipodrot.
 DR Pfam: PF01551; Peptidase M37_1
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN, PARTIAL.
 KW Inner membrane; lipoprotein.
 FT NON_TER 1
 SQ SEQUENCE 97 AA; 10667 MW; 9E194F9A13D0838E CRC64;
 Query Match Best Local Similarity 6.9%; Score 115; DB 1; Length 97;
 Matches 28; Conservative 21; Mismatches 41; Indels 2; Gaps 1;
 QY 232 SGKDDLLINASNAAGTVIGADHNDGAS--IVIOHTNGFVSSYIHKQAVKGTGTVRTGQ 289
 DB 3 AGSRQQPFATYANGVVYVAGNALGGYGNLIIKNDYLSAYVNDTWLVREGEVKGQ 62
 QY 290 RIASKNQPQSGALFEFRISRNQVVDPLTVL 321
 DB 63 KIATWGSTGTSSVRLHFEIRYKGSVNPRLVL 94
 RESULT 28
 MOZ_HUMAN STANDARD; PRT; 2004 AA.
 AC Q92794;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Monocytic leukemia zinc finger protein (Zinc finger protein 220).
 GN ZNF220 OR MOZ.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96376968; PubMed=8782817;
 RA Borow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
 RA Chaganti R.S.R., Civan C.I., Distche C., Dube I., Fritsch A.M.,
 RA Horman D., Miteiman F., Volinia S., Watmore A.E., Housman D.E.;
 RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
 RT a putative acetyltransferase to the CREB-binding protein";
 RL Nat. Genet. 14:33-41(1996).
 CC -1- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL
 CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMERA OBSERVED IN THE
 CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
 CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: BELONGS TO THE MOZ (SAS/MOZ) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U47742; AAC50662.1; -;
 DR Genew; HGNC:13013; ZNF220.
 DR MIM: 601408; -;
 DR InterPro: IPR001386; Histone_H1/H5.
 DR InterPro: IPR002717; MOZ_SAS.

01-JUL-1993 (Rel. 26, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Beta-mannanase/endo-glucanase A precursor [includes: Mannan endo-1,4-
 beta-mannanase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
 mannase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (cellulase)].
 GN MANNA
 OS Caldocellum saccharolyticum (Caldicellulosivor saccharolyticus).
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
 CC Caldicellulosivor.
 CC NCBI_Taxid=41001;
 OK
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93119139; PubMed=1476429; Bergquist P.L.;
 RA Gibbs M.D.; Saul D.J.; Luthi E.; Bergquist P.L.;
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
 RT multidomain enzyme.";
 RT Appl. Environ. Microbiol. 58:3864-3867(1992).
 RL
 RP SEQUENCE OF 1-346 FROM N.A.
 RA MEDLINE=9147819; PubMed=2039230;
 RA Luehri B.; Jasmat N.B.; Grayling R.A.; Love D.R.; Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a
 RT gene coding for a beta-mannanase from the extremely thermophilic
 RT bacterium 'Caldocellum saccharolyticum'.";
 RT Appl. Environ. Microbiol. 57:694-700(1991).
 CC
 CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT
 CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
 CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
 CC
 CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 CC linkages in mannans, galactomannans, glucomannans, and
 CC galactoglucomannans.
 CC
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC
 CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
 CC CELSIUS.
 CC
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO CELLULASE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO CELLULASE FAMILY
 CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L01257; AAA71887.1; -;
 CC EMBL: M36063; AAA72861.1; -;
 CC PIR: B43745; B43745.
 CC PIR: A48954; A48954.
 CC HSSP: O06851; INBC.
 CC InterPro: IPR001956; CBD 3.
 CC InterPro: IPR001547; GH 5.
 CC Pfam: PF00150; cellulase_1.
 CC Pfam: PF00342; CBM 3, 2.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC Hydrolyase; Glycosidase; Cellulose degradation; Signal;
 CC Multifunctional enzyme.
 CC
 CC SIGNL 1
 CC CHAIN 42 1331
 CC FT 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.
 CC FT 42 1331 CATALYTIC (MANNANASE ACTIVITY).
 CC FT 326 361 PRO/SER/THR-RICH (PT BOX).
 CC FT 362 361 SUBSTRATE-BINDING (POTENTIAL).
 CC FT 518 564 PRO/SER/THR-RICH (PT BOX).
 CC FT 564 564 SUBSTRATE-BINDING (POTENTIAL).
 CC FT 720 780 PRO/SER/THR-RICH (PT BOX).
 CC FT 721 780 CATALYTIC (ENDOGLUCANASE ACTIVITY).
 CC FT 781 1331 SIMILARITY.
 CC FT 1331 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).
 CC FT 162 257 PROTON DONOR (BY SIMILARITY).
 CC FT 257 257 NUCLEOPHILE (BY SIMILARITY).
 CC FT 338 338 T -> P (IN REF. 2).
 CC
 CC CONFLICT 338 338

FT CONFLICT 340 346 TPTPTPT -> RCHQHRQ (IN REF. 2).
 SQ SEQUENCE 1331 AA; 146892 MW; FFBGAS1BB8D8F0E0 CRC64;
 Query Match 6.8%; Score 113.5; DB 1; Length 1331;
 Best Local Similarity 23.4%; Pred. No. 2.2;
 Matches 76; Conservative 34; Mismatches 156; Indels 59; Gaps 16;
 QY 34 ASKPTVN-----STSGSGSHR-----TSGSGGLAIGSVITDSQGVPRNVYKQGDIVSK 83
 DB 630 ASNTFKKVKVTKSSVSQADYLYEIGFSGAGQLPGK---DTGEIGIQRFKSKMSVNYQ 685
 QY 84 IAKRYGLNMRKIGHNNL--NSSYITGQWLTWSDDKAKRERSIS--GVNATHPTSP 139
 DB 666 -----GNDSWISQHTSYGENEKYTAI-DGVLVWGGEPSGTTSPSTPTVTPPTPT 739
 QY 140 VAVGSSRPVQOHPAVQKPTTPPVVVKKPTTPPVVQOQPARV--APPTAPPAATGSSG 196
 DB 740 TPTPTPTPTPTPTV-TPTPTVATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 798
 QY 197 VMQFRYPVGAATPV-----VRFQTATVAGSTVTSN-----GMFSGRGGDLI----- 239
 DB 799 TQISPIYIGANQDIEGVHSARRLGKRLTGYMNNFNSNGNDWYHSSDYLCWSMGIS 858
 QY 240 --NASNAGTVIQADHNM---DGASIVIQHTNGFVSSYIHKD--AOYKTDGTVRTGQRI 291
 DB 859 GEDAKVPAAVVSKEFEYSILKNNVSAVTLQWAGYVS-----KDMYGVISENETA-PSNRW 912
 QY 292 ASMKQPSGALFEFRISRNGVYVD 316
 DB 913 AEVFKKQDAPLSLNDLINDNFVYMD 937
 RESULT 31
 HSI1 HUMAN
 ID HSI1 HUMAN STANDARD, PRT, 280 AA.
 AC Q14469;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription factor HSI-1 (Hairy and enhancer of split 1) (Hairy-
 DE like) (HHL) (Hairy homolog).
 GN HSI1 OR HRY OR HU.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 CC NCBI_Taxid=9606;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=94292187; PubMed=8020957;
 CC Feder J.N.; Li L.; Jan Y.-N.;
 CC "Genomic cloning and chromosomal localization of HRY, the human
 CC homolog to the Drosophila segmentation gene, hairy.";
 CC Genomics 20:56-61(1994).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC TISSUE=Kidney;
 CC Yao J.; Yeung S.; Sun H.; Chen N.;
 CC "Functional analysis of human HRY in Drosophila.";
 CC Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 CC [3]
 CC SEQUENCE FROM N.A.
 CC Watanabe K.; Kumagai A.; Itakura S.; Yamazaki M.; Tachiro H.; Ota T.;
 CC Suzuki Y.; Ohtsuka M.; Nishi T.; Shibahara T.; Tanaka T.;
 CC Nakamura Y.; Isono T.; Sugano S.;
 CC "NEO human cDNA sequencing project.";
 CC Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH
 CC OF MYOGENESIS BY INHIBITING THE FUNCTIONS OF MYO AND ASH1 (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
 CC WITH A CO-REPRESSOR PROTEIN (GROUCHO) (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
CC HELIX-INTERPRETING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),
CC RATHER THAN THE CANONICAL E-BOX (CANNTG).
CC -1- DOMAIN: THE CARBOXYL-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL
CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO, A
CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
CC HAIRY-RELATED PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL, L19314; AAA65220.1; -.
CC EMBL, AF264785; AAF73060.1; -.
CC EMBL, AK000415; BAA91149.1; -.
CC TRANSFAC; T04892; -.
CC Genew; HGNC:5192; HRX.
CC MIM; 139605; -.
CC InterPro; IPR001092; HLH_basic.
CC InterPro; IPR003650; Orange.
CC Pfam; PF00010; HLH_1.
CC SMART; SM00353; HLH_1.
CC SMART; SM00511; ORANGE; 1.
CC PROSITE; PS00038; HLH_1; 1.
CC PROSITE; PS50888; HLH_2; 1.
CC Transcription regulation; DNA-binding; Nuclear protein; Repressor.
CC DNA BIND 35 47 BASIC DOMAIN.
CC FT DOMAIN 48 92 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC FT DOMAIN 156 246 PRO-RICH.
CC FT DOMAIN 249 273 SER/THR-RICH.
CC FT DOMAIN 275 278 WRPW MOTIF (REQUIRED FOR ACTIVITY)
CC (BY SIMILARITY).
CC FT SEQUENCE 280 AA; 29541 MW; F9342A88FC749E3C CRC64;
SQ
Query Match 6.6%; Score 109.5; DB 1; Length 280;
Best Local Similarity 27.3%; Pred. No. 0.73;
Matches 41; Conservative 15; Mismatches 49; Indels 45; Gaps 7;
QY 126 SISSGVNT-----AHTPPVA-VQSSRPVQOHAPVOKPPVAVVKKP-----TTPV 172
DB 126 STCEVNTBEVTRLLGLHLCNCTQINAMTYEQPHPALQAPPPPGGPGHAFAPAP 185
QY 173 PVVQOPAVAPVTEAPATGS-----SGWQFRYPVGA--TNPV 211
DB 186 PLVPFPGAAPPGAPFCKLGSQAGEAAKVFQGVVAPAPDQFAFLIPNGAFHSGPVI 245
QY 212 RRF-----GTAT-----VAGSTVTSNGMW 230
DB 246 PVTYSNGTSGVGNAPNAPSSPSGSLTADSMW 275
RESULT 32
MYSC_DICDI STANDARD; PRT; 1181 AA.
ID MYSC_DICDI
AC P42522;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IC heavy chain.
GN MYC OR DMIC.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
OX NCBI_TaxID=44689;
RX NCB1_TaxID=44689;
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=95348228; PubMed=7622596;

RA Peterson M.D., Novak K.D., Reedy M.C., Ruman J.I., Titus M.A.;
RT "Molecular genetic analysis of myoc, a Dictyostelium myosin I";
RL J. Cell Sci. 108:1093-1103(1995).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- SIMILARITY: CONTRAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL, L35323; AAC37427.1; -.
CC HSSP; P08799; IMND.
CC DictyDB; DD01090; MYOC.
CC InterPro; IPR001452; SH3.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00063; myosin_head; 2.
CC PRINTS; PR00193; MYOSINHEAVY.
CC PRINTS; PR00452; SH3DOMAIN.
CC Prodom; PD000066; SH3; 1.
CC Prodom; PD000355; myosin_head; 1.
CC SMART; SM00242; MYSC; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS50002; SH3; 1.
CC Myosin; Actin-binding; ATP-binding; Chemotaxis; SH3 domain;
CC Multigene family.
CC FT DOMAIN 1 2 MYOSIN HEAD-LIKE.
CC FT NP BIND 109 116 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
CC FT DOMAIN 1122 1181 ATP (POTENTIAL).
CC FT SEQUENCE 1181 AA; 132915 MW; 5EB1BEA7F0CA8803 CRC64;
SQ
Query Match 6.6%; Score 109.5; DB 1; Length 1181;
Best Local Similarity 35.0%; Pred. No. 3.6;
Matches 36; Conservative 17; Mismatches 31; Indels 19; Gaps 6;
QY 120 LKVERSISSGVNTAHT-----PSPVAVQSS--RPVQOHAPVOKTPPVVAVVKKRP 170
DB 943 LKQOQFHASGLPASTYVAKVRKNSOVSTPSKFIAPKAPVAPVAVKPSGSGSVIMKKPAP 1001
QY 171 T-----PVVQOPAVAPVAP--PVTAPF-ATGSSGVNQFRYPV 204
DB 1002 AAPSGPPVMKKPAPFAPCGAPMVKKAPAPGAPMVKKAPAPV 1044
RESULT 33
VGP3_EBV STANDARD; PRT; 907 AA.
ID VGP3_EBV
AC P03200; P03201;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Envelope glycoprotein GP340 (Membrane antigen) (MA) [Contains:
DE Glycoprotein GP220].
GN BLF1.
OS Epstein-barr virus (strain B95-8) (human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RX NCB1_TaxID=10377;
RP SEQUENCE FROM N.A.
RC MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;

"DNA sequence and expression of the B95-8 Epstein-Barr virus genome."; Nature 310:207-211 (1984).

-!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN B-CELLS.

-!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL ENVELOPE.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: V01555; CAA24854.1; -

CC PIR: A03762; Q08E21.

CC PIR: S33008; S33008.

CC Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.

CC

CC CARBOHYD 47 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 87 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 114 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 166 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 169 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 229 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 277 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 318 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 328 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 345 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 356 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 378 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 386 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 411 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 435 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 443 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 457 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 487 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 519 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 533 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 547 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 568 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 589 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 610 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 624 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 627 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 656 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 683 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 701 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 735 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 746 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 755 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 780 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 815 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 858 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 888 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT VARSPLIC 502 MISSING (IN GP220).

CC SEQUENCE 907 AA; 94431 MW; 0750141CBGAC5C9 CRC64;

Query Match

Best Local Similarity 6.5%; Score 109; DB 1; Length 907;

Matches 66; Conservative 34; Mismatches 12; Indels 62; Gaps 13;

QY 35 SKPTVSTGSGSHRISGSGGLAISGVITDSGVNRRY--QVKQDPTSKIAQKRGGLM 92

DB 430 TSELTATGFPADNTTIG-----LPSSTHVPNLTAAPASTGPTVS----- 469

QY 93 REIGHNINNSSTYITGQMLT---LWSGDLKVRERSISGVNATHTSPVAVQSSRPP 148

DB 470 --TADVTSPTPACTTSGASFPVTPSPFWNDGTESKAPDMTSTSPVTTPTNATSPPT--- 524

QY 149 VQOHFAVOKPPTP---FVVVKKPTP--TPVVQOPAPVAPVTEADPATG-----SSGV 197

DB 525 ---PAVTPPTNATSPTPATVTPPTNATSPTLKSTPSTAVTPPTNATSPTLKSTP 580

QY 198 MQFRYPV-GATNPVRRFGTAVAGSTVTSNGMWF--GRDDLLNAGNACTVIOADHMD 255

DB 581 SAVTPPTNATSPTLGK-TSEPTSAVTPPTPATGPVGEPTSPOANATN-----HTLG 631

QY 256 GASIVIOHTNGFVSSYTHIKQAVKGTDPVATGARIAMKQPS 299

DB 632 GTS-----PTVVTSTOPKNATSAVLTGCHNTSTSSMSIRPS 670

RESULT 34

EXTN_SORBI STANDARD; PRT; 283 AA.

AC P24152.

DI 01-MAR-1992 (Rel. 21, Created)

DI 01-MAR-1992 (Rel. 21, Last sequence update)

DI 30-MAR-2000 (Rel. 39, Last annotation update)

DE Extensin precursor (Proline-rich glycoprotein).

GN HRGP.

OS Sorghum bicolor (Sorghum) (Sorghum vulgare).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;

OC Panicoideae; Andropogoneae; Sorghum.

OX NCBI_TaxID=4558;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE=Leaf;

RA MEDLINE=91370882; PubMed=1893107;

RA Raz R., Cretin C., Puigdomenech P., Martinez-Izquierdo J.A.;

RT "The sequence of a hydroxyproline-rich glycoprotein gene from sorghum vulgare.";

RT Plant Mol. Biol. 16:365-367(1991).

CC -!- FUNCTION: STRUCTURAL COMPONENT IN PRIMARY CELL WALL.

CC -!- PTM: EXTENSIN CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE

CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN GLYCOSYLATED.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL: X56010; CAA39485.1; -

CC PIR: S14449; S14449.

CC Interpro; IPR002965; P-rich extensin.

CC PRINTS; PR01217; PRICEXTNSN.

CC Repeat; Cell wall; Glycoprotein; Signal; Structural protein;

CC Hydroxylation.

CC FT SIGNAL 1 24 POTENTIAL.

CC FT CHAIN 25 283 EXTENSIN.

CC SEQUENCE 283 AA; 29593 MW; 8D7FCDD0D8ED2D90 CRC64;

Query Match

Best Local Similarity 6.5%; Score 108.5; DB 1; Length 283;

Matches 27; Conservative 7; Mismatches 23; Indels 15; Gaps 3;

QY 135 HRPSPVAVOSSRP--PVOOHFAVOKPPTPVVVVKKPTPVPVQOPAP-----VA 182

DB 117 YRSPKPRATKPTPTPTKPRATKPTPTPVLT--PSKRPVTKPPTKPRTPPVYTPMK 173

QY 183 PPTVAPATGS 194

DB 174 PPTVTKPPTHTPS 185

RESULT 35

GPI_CHURE


```

ID GPI CHLRE STANDARD; PRT; 555 AA.
AC Q9FP06; Q03927;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich
DE glycoprotein 1).
GN GPI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21159092; PubMed=11258910;
RA Ferris P.J., Moesener J.P., Waffenschmidt S., Kitz S., Drees J.,
RA Goodenough U.W.;
RT "Glycosylated polypeptide II rods-with-kinks as a structural motif in
RT plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT encoding cell wall hydroxyproline-rich glycoproteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -1- FUNCTION: Major component of the outer cell wall W6 (crystalline)
CC layer.
CC -1- SUBUNIT: Associates with GP2 and GP3.
CC -1- PTM: N-glycosylated and O-glycosylated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; AF309494; AAG45420.1; -.
DR EMBL; M58496; AAG69706.1; ALT_SEQ.
DR GlycoSuiteDB; Q9FP06; -.
DR InterPro; IPR002965; P-rich_extensin.
DR InterPro; IPR003882; Pich1_extensin.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLXTENSIN.
DR KEGG; glycoprotein; Repeat; signal.
FT SIGNAL 1 29
FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PREP. REPEATS.
FT DOMAIN 259 279 POLY-PRO.
FT CARBOHYD 359 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 555 AA; 54219 MW; 6A584A9046502F5 CRC64;

Query Match 6.5%; Score 108.5; DB 1; Length 555;
Best Local Similarity 44.6%; Pred. No. 1.8;
Matches 25; Conservative 4; Mismatches 18; Indels 9; Gaps 2;

```

```

136 TPSPVAVQSSRPVQOHPAVVQKPTPPVVVKKPTTPPVVQOQAP--VAPVTEAP 189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 154 SPSPVPVPSPPVPPVPPSPAPSPPTP-----PSPSFPVPPSPAPSPAPVPPSP 202

```

```

RESULT 36.
ID F1AB_BORPE STANDARD; PRT; 3591 AA.
AC P1235;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)

```

```

DE Filamentous hemagglutinin.
GN F1AB.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90355839; PubMed=238859;
RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
RT haemagglutinin: a protein processed from an unusually large
RT precursor.";
RL Mol. Microbiol. 4:787-800(1990).
RN [2]
RP SEQUENCE OF 1-3261 FROM N.A.
RX MEDLINE=89202384; PubMed=2539596;
RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT sequence and crucial role in adherence.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -1- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
CC INFECTON.
CC -1- SUBCELLULAR LOCATION: SURFACE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; M60351; AAA22874.1; -.
DR EMBL; M60351; AAA22875.1; ALT_INIT.
DR EMBL; M60351; AAA22976.1; ALT_INIT.
KW Antigen; Hemagglutinin.
SQ SEQUENCE 3591 AA; 367420 MW; EF7418B30D6E5138 CRC64;

Query Match 6.5%; Score 108.5; DB 1; Length 3591;
Best Local Similarity 24.5%; Pred. No. 15;
Matches 54; Conservative 21; Mismatches 74; Indels 71; Gaps 11;

```

```

6 AINSONOKPIRLGLIVITTCILAGCASKPTYNSTSGSHRTPSGGLAIG--SQVI 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3165 AVNAQNLK-----DYRDKG-----GGGNAVIGSTTL 3193

64 TDSQGVV-----NRQVKGQDTV-----SKIAQRYGLMREIGHINNINSSVITYTGWL 113
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3194 APTGVAFGRVAGBDYQAEQATIDVGOTKPALQVGGVKGTLNQDAQATV--QRN 3251

114 TLNKG--DLKVRERST--SSGVNTATPSPVAVQSS--SR-----PVQHPAVQKP 158
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3252 KHMAGGSGSEFVSAAKSKKKQVPRVETPTDVAVDGPPSRPTTPPASPQIRATVEVSSP 3311

159 TP-----PVVVKKPTPPVQOHPAVVQKPTPPVVVKKPTTPPVVQOQAPVTE 187
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3312 PPVSVATVETVPRKRVETAPLPVPPVAVQVVPVTPPKVE 3351

```

```

RESULT 37
ID A112_HUMAN STANDARD; PRT; 1593 AA.
AC P56397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
CX NCBI_Taxid=9606;
RP [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Fetal Lung;
RX MERRLIN=21264577; PubMed=11279086;
RA Cal S. Agacelles J.M., Fernandez P.L., Lopez-Ofin C.;
RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RT J. Biol. Chem. 276:17932-17940(2001)
RL J. COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. It widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL, AJ250725; CAC20419.1; -;
CC GeneW, HGNC:14605; ADAMTS12.
CC MIM, 606184; -;
DR InterPro, IPR0011762; Disintegrin.
DR InterPro, IPR002870; Pep_M12B_propep.
DR InterPro, IPR001580; Reprolysin.
DR InterPro, IPR000884; TSP1.
DR InterPro, IPR000130; Zn_MTPeptide.
DR Pfam, PF00090; tsp_1; 6.
DR Pfam, PF01421; Reprolysin; 1.
DR Pfam, PF01562; Pep_M12B_propep; 1.
DR SMART, SM00209; TSP1; 8.
DR PROSITE, PS00215; ADAM_MERPO_1; FALSE_NEG.
DR PROSITE, PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE, PS00092; TSP1; 2.
DR PROSITE, PS00142; ZINC_PROTEASE_1;
KM Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 545 596 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 CYS-RICH.
FT DOMAIN 827 881 SPACER 1.
FT DOMAIN 886 943 TSP TYPE-1 2.
FT DOMAIN 947 995 TSP TYPE-1 3.
FT DOMAIN 996 1315 TSP TYPE-1 4.
FT DOMAIN 1316 1364 SPACER 2.
FT DOMAIN 1367 1423 TSP TYPE-1 5.
FT DOMAIN 1426 1471 TSP TYPE-1 6.
FT DOMAIN 1472 1571 TSP TYPE-1 7.
FT DOMAIN 1572 1671 TSP TYPE-1 8.
FT DOMAIN 1672 1771 TSP TYPE-1 9.
FT DOMAIN 1772 1871 TSP TYPE-1 10.
FT DOMAIN 1872 1971 TSP TYPE-1 11.
FT DOMAIN 1972 2071 TSP TYPE-1 12.
FT DOMAIN 2072 2171 TSP TYPE-1 13.
FT DOMAIN 2172 2271 TSP TYPE-1 14.
FT DOMAIN 2272 2371 TSP TYPE-1 15.
FT DOMAIN 2372 2471 TSP TYPE-1 16.
FT DOMAIN 2472 2571 TSP TYPE-1 17.
FT DOMAIN 2572 2671 TSP TYPE-1 18.
FT DOMAIN 2672 2771 TSP TYPE-1 19.
FT DOMAIN 2772 2871 TSP TYPE-1 20.
FT DOMAIN 2872 2971 TSP TYPE-1 21.
FT DOMAIN 2972 3071 TSP TYPE-1 22.
FT DOMAIN 3072 3171 TSP TYPE-1 23.
FT DOMAIN 3172 3271 TSP TYPE-1 24.
FT DOMAIN 3272 3371 TSP TYPE-1 25.
FT DOMAIN 3372 3471 TSP TYPE-1 26.
FT DOMAIN 3472 3571 TSP TYPE-1 27.
FT DOMAIN 3572 3671 TSP TYPE-1 28.
FT DOMAIN 3672 3771 TSP TYPE-1 29.
FT DOMAIN 3772 3871 TSP TYPE-1 30.
FT DOMAIN 3872 3971 TSP TYPE-1 31.
FT DOMAIN 3972 4071 TSP TYPE-1 32.
FT DOMAIN 4072 4171 TSP TYPE-1 33.
FT DOMAIN 4172 4271 TSP TYPE-1 34.
FT DOMAIN 4272 4371 TSP TYPE-1 35.
FT DOMAIN 4372 4471 TSP TYPE-1 36.
FT DOMAIN 4472 4571 TSP TYPE-1 37.
FT DOMAIN 4572 4671 TSP TYPE-1 38.
FT DOMAIN 4672 4771 TSP TYPE-1 39.
FT DOMAIN 4772 4871 TSP TYPE-1 40.
FT DOMAIN 4872 4971 TSP TYPE-1 41.
FT DOMAIN 4972 5071 TSP TYPE-1 42.
FT DOMAIN 5072 5171 TSP TYPE-1 43.
FT DOMAIN 5172 5271 TSP TYPE-1 44.
FT DOMAIN 5272 5371 TSP TYPE-1 45.
FT DOMAIN 5372 5471 TSP TYPE-1 46.
FT DOMAIN 5472 5571 TSP TYPE-1 47.
FT DOMAIN 5572 5671 TSP TYPE-1 48.
FT DOMAIN 5672 5771 TSP TYPE-1 49.
FT DOMAIN 5772 5871 TSP TYPE-1 50.
FT DOMAIN 5872 5971 TSP TYPE-1 51.
FT DOMAIN 5972 6071 TSP TYPE-1 52.
FT DOMAIN 6072 6171 TSP TYPE-1 53.
FT DOMAIN 6172 6271 TSP TYPE-1 54.
FT DOMAIN 6272 6371 TSP TYPE-1 55.
FT DOMAIN 6372 6471 TSP TYPE-1 56.
FT DOMAIN 6472 6571 TSP TYPE-1 57.
FT DOMAIN 6572 6671 TSP TYPE-1 58.
FT DOMAIN 6672 6771 TSP TYPE-1 59.
FT DOMAIN 6772 6871 TSP TYPE-1 60.
FT DOMAIN 6872 6971 TSP TYPE-1 61.
FT DOMAIN 6972 7071 TSP TYPE-1 62.
FT DOMAIN 7072 7171 TSP TYPE-1 63.
FT DOMAIN 7172 7271 TSP TYPE-1 64.
FT DOMAIN 7272 7371 TSP TYPE-1 65.
FT DOMAIN 7372 7471 TSP TYPE-1 66.
FT DOMAIN 7472 7571 TSP TYPE-1 67.
FT DOMAIN 7572 7671 TSP TYPE-1 68.
FT DOMAIN 7672 7771 TSP TYPE-1 69.
FT DOMAIN 7772 7871 TSP TYPE-1 70.
FT DOMAIN 7872 7971 TSP TYPE-1 71.
FT DOMAIN 7972 8071 TSP TYPE-1 72.
FT DOMAIN 8072 8171 TSP TYPE-1 73.
FT DOMAIN 8172 8271 TSP TYPE-1 74.
FT DOMAIN 8272 8371 TSP TYPE-1 75.
FT DOMAIN 8372 8471 TSP TYPE-1 76.
FT DOMAIN 8472 8571 TSP TYPE-1 77.
FT DOMAIN 8572 8671 TSP TYPE-1 78.
FT DOMAIN 8672 8771 TSP TYPE-1 79.
FT DOMAIN 8772 8871 TSP TYPE-1 80.
FT DOMAIN 8872 8971 TSP TYPE-1 81.
FT DOMAIN 8972 9071 TSP TYPE-1 82.
FT DOMAIN 9072 9171 TSP TYPE-1 83.
FT DOMAIN 9172 9271 TSP TYPE-1 84.
FT DOMAIN 9272 9371 TSP TYPE-1 85.
FT DOMAIN 9372 9471 TSP TYPE-1 86.
FT DOMAIN 9472 9571 TSP TYPE-1 87.
FT DOMAIN 9572 9671 TSP TYPE-1 88.
FT DOMAIN 9672 9771 TSP TYPE-1 89.
FT DOMAIN 9772 9871 TSP TYPE-1 90.
FT DOMAIN 9872 9971 TSP TYPE-1 91.
FT DOMAIN 9972 10071 TSP TYPE-1 92.
FT DOMAIN 10072 10171 TSP TYPE-1 93.
FT DOMAIN 10172 10271 TSP TYPE-1 94.
FT DOMAIN 10272 10371 TSP TYPE-1 95.
FT DOMAIN 10372 10471 TSP TYPE-1 96.
FT DOMAIN 10472 10571 TSP TYPE-1 97.
FT DOMAIN 10572 10671 TSP TYPE-1 98.
FT DOMAIN 10672 10771 TSP TYPE-1 99.
FT DOMAIN 10772 10871 TSP TYPE-1 100.
FT DOMAIN 10872 10971 TSP TYPE-1 101.
FT DOMAIN 10972 11071 TSP TYPE-1 102.
FT DOMAIN 11072 11171 TSP TYPE-1 103.
FT DOMAIN 11172 11271 TSP TYPE-1 104.
FT DOMAIN 11272 11371 TSP TYPE-1 105.
FT DOMAIN 11372 11471 TSP TYPE-1 106.
FT DOMAIN 11472 11571 TSP TYPE-1 107.
FT DOMAIN 11572 11671 TSP TYPE-1 108.
FT DOMAIN 11672 11771 TSP TYPE-1 109.
FT DOMAIN 11772 11871 TSP TYPE-1 110.
FT DOMAIN 11872 11971 TSP TYPE-1 111.
FT DOMAIN 11972 12071 TSP TYPE-1 112.
FT DOMAIN 12072 12171 TSP TYPE-1 113.
FT DOMAIN 12172 12271 TSP TYPE-1 114.
FT DOMAIN 12272 12371 TSP TYPE-1 115.
FT DOMAIN 1237

FT	METAL	392	392	ZINC (CATALYTIC) (BY SIMILARITY) .
FT	ACT SITE	393		ZINC (CATALYTIC) (BY SIMILARITY) .
FT	METAL	396	396	ZINC (CATALYTIC) (BY SIMILARITY) .
FT	METAL	402	402	ZINC (CATALYTIC) (BY SIMILARITY) .
FT	CARBOHYD	105	105	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	125	125	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	215	215	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	485	485	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	685	685	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	790	790	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	951	951	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	1104	1104	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	1275	1275	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	1300	1300	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	1371	1371	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	1503	1503	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	SEQUENCE	1593 AA:	17545 MW;	079F48E5B83A3 CRC64;

Query Match	6.5%;	Score 108;	DB 1;	Length 1593;	
Best Local Similarity	21.2%;	Pred. No. 6.4;			
Matches	57;	Conservative	35;	Mismatches 93; Indels 84; Gaps 10;	
Qy	80	TVSTIAQRGLNMEIGHINNLSNTYITVGQW----	LTLSWGDGLKXERGISG----	129	
Db	1054	TASKEGDLGGQWQDSSSTQPELSRYYLISGTSQPLTSGSLISQSEENVSSSDTPT		1113	
Qy	130	--GNVTAHTEBPVAVOSSRPV-----	QCHPAVKPFP	161	
Db	1114	SEGGLVATTTSGSSLSRBNITPVPFPYNTLTGKPEMEIHSGSGEERQPEDKDSNP		1173	
Qy	162	VV--VVKKPTTPPVQQAPVAPVPE-----	APFAIGSSGVMQFRFPVAGTNV	211	
Db	1174	VITMKIRPFGDAIVESTEMPLAPLPPLDLSRESWMPFFSVYMGDLR-----	SQRTT	1227	
Qy	212	RRFGTAIVAGSTVTSNGMWFSGRDODLINSAGTVLO--	ADHNMDDASLYIQHTNGVVS	269	
Db	1228	SETGPRVEG-----	WTEKPAINTLPLGSDHOPE-----	PSGRTA	1265
Qy	270	SYIHK-----DAQVKGDTVTRQRIASM		294	
Db	1264	NRNHLKLPNNNQTSSEPLVLEEDATSL		1292	

ID	ALYS_ENTFPA	STANDARD;	PRT;	671 AA.
ALYS_38	RESULT 18			
AC	P37710;			
DN	01-OCT-1994 (Rel. 30. Created)			
DN	01-OCT-1994 (Rel. 30. Last sequence update)			
DT	16-OCT-2001 (Rel. 40. Last annotation update)			
DE	Autolysin precursor (EC 3.2.1.-) (Peptidoglycan hydrolase)			
DE	(Beta-glycosidase).			
OS	Enterococcus faecalis (Streptococcus faecalis)			
OC	Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.			
OX	NCBI_TaxId=1351;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91358349; PubMed=1679432;			
RA	Beliveau C., Potvin C., Trudel J., Asselin A., Bellemare G.;			
RT	"Cloning, sequencing, and expression in Escherichia coli of a			
RT	Streptococcus faecalis autolysin".			
RL	J. Bacteriol. 173:5619-5623 (1991).			
CC	- FUNCTION: HYDROLYZES THE CELL WALL OF E.FAECALIS AND			
CC	M. LYSOGENICTICS. MAY PLAY AN IMPORTANT ROLE IN CELL WALL GROWTH			
CC	AND CELL SEPARATION.			
CC	- SUBCELLULAR LOCATION: Secreted (Probable).			
CC	- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN			
CC	BINDING.			
CC	- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.			
CC	- SIMILARITY: CONTAINS 5 LYSM REPEATS.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M58002; AAA67325.1; -
 DR PIR; A38109; A38109.
 DR InterPro; IPR002901; Amidase_4.
 DR InterPro; IPR002482; LYSM.
 DR Pfam; PF01476; LYSM; 5.
 DR Pfam; PF01832; Amidase_4; 1.
 DR SMART; SMO0047; LY22; 1.
 DR SMART; SMO0257; LYSM; 5.
 DR Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
 KM Cell division; Septation; Repeat; Signal.
 FT SIGNAL
 FT CHAIN 1 53 POTENTIAL.
 FT REPEAT 363 405 LYSM 1.
 FT REPEAT 431 473 LYSM 2.
 FT REPEAT 499 541 LYSM 3.
 FT REPEAT 567 609 LYSM 4.
 FT REPEAT 629 671 LYSM 5.
 SQ SEQUENCE 671 AA; 70442 MW; 34582686C6C1A4A33 CRC64;

Query Match 6.4%; Score 107.5; DB 1; Length 671;
 Best Local Similarity 20.8%; Pred. No. 2.6;
 Matches 66; Conservative 43; Mismatches 77; Indels 131; Gaps 15;

QY 34 ASKPTVNS-----TSGSGSHRTSGSGGALISQVITDSQGVNRYQVQOG 78
 DB 311 ATDPSYNAKANNVITAYNLGYDITSSGSGN--TGGCTYNPRTGSGNNSGNTTYTVVSG 368
 QY 79 DTVSKIAQRYGL-----NMREI-----GHINNLSY 105
 DB 369 DTLNKIAQYGVSVANLRSNMGISGDLIFVGQKLIIVKKASGNTGSGNNSGNTN 428
 QY 106 TIYTGQMTLWSGDLKXVERSSGVNTAHTPSPYAVOSSRPVQGHAVQKPPPVVVV 165
 DB 429 TTY-----TVKSGD--TLNKIAQYGVTVANLRSNMGISGDLIFVGQK-----LIV 472
 QY 166 KKPFTPPVQGPAPVAPVTEAPATGSSGVNQPRYPVVGATNPVRRFGATVAGSTVT 225
 DB 473 KKGFT-----SGNTG-----GSSNN 491
 QY 226 SNG--MMFSGRDGLINASNAGTVIQADNMDGASIVIQHTNGFVSSYHIKDAQVKTGD 283
 DB 492 QSGNTVYTVTIKSGDTLTK-----IAQYGVSVAN--LRSNMGI-----SGD 530
 QY 284 TVRTGQRIASMKQPSG 300
 DB 531 LIFAGQKLI-VKGTSG 546

RESULT 39
 FRZE_MYXA
 ID FRZE_MYXA STANDARD; PRT; 777 AA.
 AC P18769;
 DT 01-NOV-1990 (Rel. 16; Created)
 DT 01-NOV-1990 (Rel. 16; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE Gliding motility regulatory protein (EC 2.7.3.-).
 GN FRZE.
 OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 OC Myxococcales; Cyetobacterineae; Myxococcaceae; Myxococcus.
 CX NCBI_TaxId=34;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90332690; PubMed=2165608;
 RA McCleary W.R.; Zusman D.R.;

RT "Frze of Myxococcus xanthus is homologous to both Chea and Chey of
 RT Salmonella typhimurium."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5898-5902 (1990).
 RN [2]
 RP PHOSPHORYLATION OF HIS-49.
 RX MEDLINE=91072208; PubMed=2123853;
 RA McCleary W.R.; Zusman D.R.;
 RT "Purification and characterization of the Myxococcus xanthus Frze
 RT protein shows that it has autophosphorylation activity."
 RL J. Bacteriol. 172:6661-6668 (1990).
 CC -1- FUNCTION: FRZE IS INVOLVED IN A SENSORY TRANSDUCTION PATHWAY THAT
 CC CONTROLS THE FREQUENCY AT WHICH CELLS REVERSE THEIR GLIDING
 CC DIRECTION. FRZE SEEMS TO BE CAPABLE OF AUTOPHOSPHORYLATING ITSELF
 CC ON AN HISTIDINE RESIDUE AND THEN TO TRANSFER THAT GROUP TO AN
 CC ASPARTATE RESIDUE IN THE C-TERMINAL PART OF THE PROTEIN.
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -1- SIMILARITY: SIMILAR TO BOTH CHEA AND CHEY.
 CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M35192; AAA25396.1; -
 DR PIR; A35966; A35966.
 DR HSSP; Q56310; 1B30.
 DR InterPro; IPR003594; ATPbind_Atpase.
 DR InterPro; IPR004358; Bact_sense_pr_C.
 DR InterPro; IPR002455; Chew.
 DR InterPro; IPR004359; HIS_KIN_sig.
 DR InterPro; IPR002570; Hpt.
 DR InterPro; IPR001789; Response_reg.
 DR Pfam; PF00072; response_reg; 1.
 DR Pfam; PF01584; Chew; 1.
 DR Pfam; PF01627; Hpt; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR PRINTS; PR00344; BCTRLENSOR.
 DR ProDom; PD000039; Response_reg; 1.
 DR ProDom; PD003142; Hpt; 1.
 DR SMART; SMO0260; Chew; 1.
 DR SMART; SMO0387; HATPase_C; 1.
 DR SMART; SMO0073; HPT; 1.
 DR SMART; SMO0448; REC; 1.
 DR PROSITE; PS50109; HIS_KIN; 1.
 DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
 KW Sensory transduction; Transfeferase; Kinase; Phosphorylation.
 FT DOMAIN 270 509 HISTIDINE KINASE.
 FT DOMAIN 660 776 RESPONSE REGULATORY.
 FT MOD RES 49 49 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT DOMAIN 130 197 ALA/PRO-RICH (POSSIBLE HINGE REGION).
 SQ SEQUENCE 777 AA; 83189 MW; 9912BD40991C69E5 CRC64;

Query Match 6.4%; Score 107.5; DB 1; Length 777;
 Best Local Similarity 23.7%; Pred. No. 3.1;
 Matches 53; Conservative 30; Mismatches 76; Indels 65; Gaps 11;

QY 110 GQMTLWSGDLKXVERSSGVNT-----AHTPSVA-----VQSRPP--VQGHFA 154
 DB 80 GKRTVEATDVLLACDVLSTLNDLGGANTGNPASEMVMALAEVSGQTPTAAGARPV 139
 QY 155 VQKPTPPVVVKKPTPPPVVQGP-----AVAPVTEAPPA--TGSSGVNQPRYPVVGATN 208
 DB 140 APPAPPPAPVAAVVPVPAVAPVQAPVAPPPQAPVAPBGAHAAAAPPA-- 197
 QY 209 PVVRRFGATVAGSTVNSNGMFGSGRDGLINASNAGTVIQADNMDGASIVIQHTNGFV 268
 DB 198 -----HGRDEAPSAKSAV--ADRSI--RVNVEVDALGLT 230
 QY 269 SSYHIKDAQVKTGDTVTRGQRIASMKQPSGALFFFRISRRG 312

Db 231 AGDLIVESAR-----GRNLSRTE---ALFE-RFRRLG 259

RESULT 40
SHK1_RAT STANDARD; PRT; 2167 AA.
AC Q9WV48; Q9WU13; Q9WU8; Q9QZ28;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP
DE interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
DE interacting protein) (SSTR interacting protein) (SSTRIP).
GN SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RN [1] SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
RN [1] DLG4.
RN [1] TISSUE=Brain;
RN [1] MEDLINE=994119021; PubMed=10488079;
RN [1] Yao I., Hata Y., Hirono K., Desnuchi M., Ide N., Takeuchi M., Takai Y.;
RN [1] "Synamon", a novel neuronal protein interacting with synapse-associated
RN [1] protein 90/postsynaptic density-95-associated protein.";
RN [1] J. Biol. Chem. 274:27463-27466(1999).
RN [1] [2]
RN [1] SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RN [1] STRAIN=Sprague-Dawley;
RN [1] MEDLINE=99360550; PubMed=10433268;
RN [1] Naibidit S., Kim E., Tu D.C., Xiao B., Sala C., Valtchanoff J.,
RN [1] Weiberg R.J., Worley P.F., Sheng M.;
RN [1] "Shank", a novel family of postsynaptic density proteins that binds to
RN [1] the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RN [1] Neuron 23:569-582(1999).
RN [1] [3]
RN [1] SEQUENCE FROM N.A. (ISOFORM 1).
RN [1] TISSUE=Brain;
RN [1] MEDLINE=20549637; PubMed=10958799;
RN [1] Tobben S., Suedhof T.C., Stahl B.;
RN [1] "The G protein-coupled receptor Ctl1 interacts directly with proteins
RN [1] of the Shank family.";
RN [1] J. Biol. Chem. 275:36204-36210(2000).
RN [1] [4]
RN [1] PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
RN [1] DEVELOPMENTAL STAGE.
RN [1] TISSUE=Brain;
RN [1] MEDLINE=99436166; PubMed=10506216;
RN [1] Lim S., Naibidit S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
RN [1] "Characterization of the Shank family of synaptic proteins. Multiple
RN [1] genes, alternative splicing, and differential expression in brain and
RN [1] development.";
RN [1] J. Biol. Chem. 274:29510-29518(1999).
RN [1] [5]
RN [1] PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RN [1] TISSUE=Brain;
RN [1] MEDLINE=20020275; PubMed=10551867;
RN [1] Zitter H., Hoernck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RN [1] "Somatostatin receptor interacting protein defines a novel family of
RN [1] multidomain proteins present in human and rodent brain.";
RN [1] J. Biol. Chem. 274:32997-33001(1999).
RN [1] [6]
RN [1] INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
RN [1] PubMed=10433269;
RN [1] Tu J.C., Xiao B., Naibidit S., Yuan J.P., Petralia R.S., Brakeman P.,
RN [1] Doan A., Aakalu V.K., Lananah A.A., Sheng M., Worley P.F.;
RN [1] "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
RN [1] postsynaptic density proteins.";
RN [1] Neuron 23:583-592(1999).
RN [1] [7]
RN [1] INTERACTION WITH SPTAN1.

CC PubMed=11509555;
RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,
RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;
RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the
RT multidomain Shank protein family interact with the cytoskeletal
RT protein alpha-fodrin.";
RL J. Biol. Chem. 276:40104-40112(2001).
RN [6]
RN [6] FUNCTION.
RN [6] PubMed=11498055;
RA Sala C., Piench V., Wilson N.R., Passafaro M., Liu G., Sheng M.;
RT "Regulation of dendritic spine morphology and synaptic function by
RT Shank and Homer.";
RL Neuron 31:115-130(2001).
RN [9]
RN [9] REVIEW.
RN [9] PubMed=10806096;
RA Sheng M., Kim E.;
RT "The Shank family of scaffold proteins.";
RL J. Cell Sci. 113:1851-1856(2000).
CC [1]
CC [1] FUNCTION: Seems to be an adapter protein in the postsynaptic
CC density (PSD) of excitatory synapses that interconnects receptors
CC of the postsynaptic membrane including NMDA-type and metabotropic
CC glutamate receptors, and the actin-based cytoskeleton. May play a
CC role in the structural and functional organization of the
CC dendritic spine and synaptic junction. Overexpression promotes
CC maturation of dendritic spines and the enlargement of spine heads
CC via its ability to recruit Homer to postsynaptic sites, and
CC enhances presynaptic function.
CC [1] SUBUNIT: May homomultimerize via its SAM domain. Interacts with
CC SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with
CC DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via
CC the PDZ domain (by similarity).
CC [1] SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
CC neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.
CC [1] ALTERNATIVE PRODUCTS: At least 5 isoforms; 1, (shown here); 2, 3,
CC 4/4 and 5, are produced by alternative splicing.
CC [1] TISSUE SPECIFICITY: Expressed only in brain (neocortical of cortex,
CC CA1 region hippocampus and molecular layer of cerebellum).
CC [1] DEVELOPMENTAL STAGE: Expression increases from low levels at birth
CC to high levels at 3-4 weeks before dropping slightly in adulthood.
CC Expressed in the cortex and the molecular layer of the cerebellum
CC at postnatal day 7. Isoform 2 expression does not change during
CC development of both cortex and cerebellum. Isoform 4 expression
CC decreases significantly during development of cortex but not
CC cerebellum.
CC [1] SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC [1] SIMILARITY: CONTAINS 7 ANK REPEATS.
CC [1] SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
CC [1] SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC [1] SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC [1] SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC [1] This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC [1] EMBL, AF102855; AAD04569.2; -
CC EMBL, AF131951; AAD29417.1; ALT_INIT.
CC EMBL, AF159046; AAD42975.1; -
CC EMBL, AF141904; AAF02498.1; ALT_INIT.
CC HSSP, P00519; IABL.
CC InterPro, IPR002110; ANK.
CC InterPro, IPR001478; PDZ.
CC InterPro, IPR001660; SAM.
CC InterPro, IPR001452; SH3.
CC Pfam, PF00023; ank; 7.
CC Pfam, PF00595; PDZ; 1.
CC Pfam, PF00536; SAM; 1.
CC Pfam, PF00018; SH3; 1.

Job time : 25 BECS

DR	ProDom.	PD000066;	SH3; 1.
DR	SMART;	SM00248;	ANK; 3.
DR	SMART;	SM00228;	PDZ; 1.
DR	SMART;	SM00454;	SAM; 1.
DR	SMART;	SM00326;	SH3; 1.
DR	PROSITE;	PS50088;	ANK_REPEAT; 3.
DR	PROSITE;	PS50297;	ANK_REP_REGION; 1.
DR	PROSITE;	PS50106;	PDZ; 1.
DR	PROSITE;	PS50002;	SH3; 1.
KW	ANK repeat;	SH3 domain;	Repeat; Alternative splicing.
FT	REPEAT	195	210
FT	REPEAT	212	245
FT	REPEAT	246	278
FT	REPEAT	279	312
FT	REPEAT	313	345
FT	REPEAT	346	378
FT	REPEAT	379	395
FT	DOMAIN	554	613
FT	DOMAIN	663	757
FT	DOMAIN	2104	2167
FT	DOMAIN	329	332
FT	DOMAIN	1010	1015
FT	DOMAIN	1022	1027
FT	DOMAIN	1194	1199
FT	DOMAIN	1850	1860
FT	VARSPLIC	1	613
FT	VARSPLIC	615	654
FT	VARSPLIC	646	654
FT	VARSPLIC	797	804
FT	VARSPLIC	1330	1243
FT	VARSPLIC	1944	2167
FT	CONFLICT	1141	1141
FT	CONFLICT	1174	1174
FT	CONFLICT	1246	1246
FT	CONFLICT	1323	1323
FT	CONFLICT	1331	1331
FT	CONFLICT	1726	1726
SO	SEQUENCE	2167 AA;	226333 MW; 3F478B5A7B18BA86 CRC64;
<hr/>			
Query Match			
Best local Similarity	6.4%;	Score 107.5;	DB 1,
Matches 69;	Conservative 43;	Pred. No. 9.8;	Length 2167;
		Mismatches 109;	Indels 81;
		Gaps	16

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 8, 2003, 10:58:52 ; Search time 80 seconds
(without alignments)
829,340 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670
Sequence: 1 MVTYIAINSQNKPIKRLGL.....LFEFRISRGVYVPLTVLK 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	268	16.0	314	2	Q9JP90
2	253	15.1	242	16	Q9P223
3	249.5	14.9	467	16	Q9CCK1
4	238	14.3	259	16	Q8X2T2
5	234	14.0	231	16	Q9HUP0
6	232.5	13.9	250	16	Q8Z3Y0
7	232.5	13.9	252	16	Q8ZM83
8	228.5	13.7	269	2	Q9L528
9	227.5	13.6	230	2	Q9K120
10	226.5	13.6	244	2	Q9MNV6
11	221	13.2	217	2	Q9JAG4
12	220	13.2	268	16	Q8Y038
13	219.5	13.1	454	16	Q9S336
14	218.5	13.1	333	16	Q8ZBQ1
15	208.5	12.5	311	16	Q9K17
16	205	12.3	427	16	Q8YGS9

17	203	12.2	415	16	Q9JTP1	Q9JTP1 neisseria m
18	202.5	12.1	379	16	Q8X7Z3	Q8X7Z3 escherichia
19	198	11.9	415	16	Q9JYP9	Q9JYP9 neisseria m
20	196.5	11.8	515	16	Q9BMD1	Q9BMD1 rhizobium l
21	196	11.7	609	16	Q9AET7	Q9AET7 caulobacter
22	188.5	11.3	562	16	Q8UBQ5	Q8UBQ5 agrobacteri
23	174.5	10.4	512	2	Q08251	Q08251 rhizobium m
24	174	10.4	512	16	Q926D2	Q926D2 rhizobium m
25	172.5	10.3	392	2	Q8RNB5	Q8RNB5 bartonella
26	171.5	10.3	312	16	Q9X7W8	Q9X7W8 streptomyce
27	171	10.2	449	16	Q8R6U3	Q8R6U3 thermococci
28	169	10.1	223	16	Q8X3L6	Q8X3L6 escherichia
29	167	10.0	371	16	Q83190	Q83190 treponema p
30	165.5	9.9	401	2	Q9KJW8	Q9KJW8 bartonella
31	158.5	9.5	436	16	Q8YAE2	Q8YAE2 listeria m
32	158	9.5	321	16	Q8RG41	Q8RG41 fusobacteri
33	158	9.5	375	16	Q9RM21	Q9RM21 deinococcus
34	155	9.3	437	16	Q9Z7Y9	Q9Z7Y9 listeria in
35	153.5	9.2	176	2	Q52606	Q52606 raietonia s
36	153	9.2	228	16	Q9X8M5	Q9X8M5 streptomyce
37	152.5	9.1	203	2	Q9X6S4	Q9X6S4 vibrio para
38	152	9.1	760	16	Q8YRU0	Q8YRU0 anabaena sp
39	147.5	8.8	176	2	Q9F7X9	Q9F7X9 escherichia
40	145	8.7	247	2	Q9S4T2	Q9S4T2 legionella
41	141	8.4	162	2	Q9R3D9	Q9R3D9 escherichia
42	141	8.4	164	2	Q9APJ3	Q9APJ3 escherichia
43	139.5	8.4	512	2	Q03491	Q03491 listeria gr
44	138.5	8.3	715	16	P74517	P74517 synchocyst
45	136.5	8.2	285	16	Q9PEP1	Q9PEP1 xylella fab

ALIGNMENTS

RESULT 1
Q9JP90 PRELIMINARY; PRT; 314 AA.
ID Q9JP90
AC Q9JP90;
DT 01-OCT-2000 (TREMURel. 15, Created)
DT 01-OCT-2000 (TREMURel. 15, Last sequence update)
DT 01-JUN-2001 (TREMURel. 17, Last annotation update)
DE N1PD protein.
GN N1PD.
OS Rhodococcus gelatinosus (Rhodopseudomonas gelatinosa).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Rubrivivax.
OX NCBI_TaxID=28068;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=IL144;
RC MEDLINE=94132007; PubMed=8300574;
RA Nagashima K.V., Shimada K., Ohama S., Shimada K.;
RT "Phylogenetic analysis of photosynthetic genes of Rhodococcus
RT gelatinosus: Possibility of horizontal gene transfer in purple
RT bacteria.";
RL Photosyn. Res. 36:185-191(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=IL144;
RC MEDLINE=94132007; PubMed=8300574;
RA Nagashima K.V., Matsura K., Ohama S., Shimada K.;
RT "Primary structure and transcription of genes encoding B870 and
RT photosynthetic reaction center apoproteins from Rubrivivax
RT gelatinosus.";
RL J. Biol. Chem. 269:2477-2484(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=IL144;
RC Igarashi N., Shimada K., Matsura K., Nagashima K.V.;
RT "Photosynthetic gene cluster in purple bacterium, Rubrivivax
RT gelatinosus." (eds.);
RL (in) Garab G. (eds.);
RT Photosynthesis:
RL mechanisms and effects (Proceedings of the 11th international congress

[1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=PW70;
 RC MEDLINE=2145866; PubMed=11248100;
 RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whitam T.S., Kapur V.,
 RT "Complete genome sequence of *Pasteurella multocida* Fm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL; AE006198; AA003698.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 2.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 467 AA; 49735 MW; A9E8CDE70DD284A6 CRC64;

Query Match 14.9%; Score 249.5; DB 16; Length 467;
 Best Local Similarity 30.0%; Pred. No. 3.6e-10;
 Matches 80; Conservative 42; Mismatches 110; Indels 35; Gaps 9;

QY 73 YQYKQGVTSKIAQRYGLNREIGHINNNLSYTIYTGWLTMSGDLKVEREISGSVN 132
 DB 215 YTRKGDITMLIAYISGLDKELASLNNMSEPRILSVGQTLRVSG---RVASTSQPV 270
 QY 133 TAHTPSVAVQSSRPVQOHPAVOKPTP-----VVVYKPTPTPPVVOQAPVA 182
 DB 271 T---QVTVPVSGP--KSESVTYTPGPHGQYGSDDTIIIGIKSVSAAPVPVQPEPV 324
 QY 183 PPV--TEAPATGSSGM---QFRYPVGAITNPVRFGRATVAGSTVTSNGMFGSRDG 236
 DB 325 KPVESTSVPVSTSSKIMVSNVTWQMPKGN--IVQGFSTADGG---NKGIDIASRG 377
 QY 237 DLINASNAIVIAQADHMDGAS--IVQHTNGFVSVYIHAKDAQVKTGDTYRTGQRIASM 294
 DB 378 QAVNAAAGVAVVAGNLRGVGNLITKENDYLSAIVANESLIVKQGEVRAQQIAYK 437
 QY 295 KNQPSGAALPEFRISRGVYVDPPLTVL 321
 DB 438 GSSGTSNVKLFELIRYKGSVDPTRYL 464

RESULT 4
 Q8X2T2 PRELIMINARY; PRT; 259 AA.
 ID Q8X2T2
 AC Q8X2T2;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative lipoprotein.
 GN ECS3738.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=2156231; PubMed=11258796;
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AP002563; BA837161.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 KM Lipoprotein.
 SQ SEQUENCE 259 AA; 27528 MW; 8A1B89E01F9B72E8 CRC64;

Query Match 14.3%; Score 238; DB 16; Length 259;
 Best Local Similarity 24.9%; Pred. No. 1.1e-09;
 Matches 78; Conservative 46; Mismatches 103; Indels 86; Gaps 8;

QY 16 KRLGLFGVITTCILACGASKPPTNYSGGSHRTSSGGGLAGSQVITTSQGVPNRYOV 75
 DB 17 KSLIGVLLSVGLLGLAGC-----SGSKSSPTGYSG-----SVYTV 52
 QY 76 KQDITVSKIAQRYGLNREIGHINNNLSYTIYTGWLTMSGDLK----VREISISGV 131
 DB 53 KRGDITLRISRTTGTSTVKEALRLNGISPPYTIIEVQRLKI--GSAKSSSTRKSTASTT 110
 QY 132 NTAH--TPSPVAVQSSRPVQOHPAVOKPTPVPVVKKPTPTPPVVOQAPVAPVTEAPF 190
 DB 111 KTASVTPSSAVPSSSMPVPVQRCMLMPTTGKILM-----PY 146
 QY 191 ATGSSGVMQFRYPVGAITNPVRFGRATVAGSTVTSNGMFGSRDGLINASNAIVIAQ 250
 DB 147 STADGG-----NKGIDIASAPRGTPVYAAAGAKVYV 177
 QY 251 DHNMDGAS--IVQHTNGFVSVYIHAKDAQVKTGDTYRTGQRIASMKNQPSGALPEFRI 308
 DB 178 GNQLRGYNLIMTKHSEDIYTAIAHNDTLPVNNQGSYKAOQKLTATGSTDPAASVRLHFOI 237
 QY 309 SRNGVYVDPPLTVL 321
 DB 238 RYRATADPLRYL 250

RESULT 5
 Q9HUP0 PRELIMINARY; PRT; 231 AA.
 ID Q9HUP0
 AC Q9HUP0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein PA4924.
 GN PA4924.
 OS *Pseudomonas aeruginosa*.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC *Pseudomonas*.
 NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAOI;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer U., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004905; AA008309.1; -
 DR HSP; P23931; IEOG.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 231 AA; 25174 MW; 3FC8EAB82A42B91 CRC64;

Query Match 14.0%; Score 234; DB 16; Length 231;
 Best Local Similarity 26.7%; Pred. No. 1.9e-09;
 Matches 68; Conservative 37; Mismatches 90; Indels 60; Gaps 6;

QY 69 VPKRYQKQDITVSKIAQRYGLNREIGHINNNLSYTIYTGWLTMSGDLKVEREISIS 128
 DB 32 VRGIVYKRGDITLYSIATRHGMVYKDLARANGIRPVAIVVQ--VAFDGRKSTYVASSR 90

QY 129 SGVNTAHTSPVAVQSSRPVQOHPAVOKPFPVAVVKKPPPPVYVQAPAPVPTEA 188
 Db 91 SSSNTR-----ARKPPPPPSVT-----108
 QY 189 PPAATSSGAWQFRYPVAGATNPVPRFGATVAGSTVNSGMMFSGRGDDLINASAGTVI 248
 Db 109 -----LKGWQPFM--KGPVIRF-----SSDDKLNKIRIAGTIGQPVQASLAKKV 153
 QY 249 QADHNDGAS--IVYQHTNGFVSSYIHKDAQVKTGDTVTRTGQRIASMKQPSGALPEF 306
 Db 154 PAVNNMGYGNLVITIQHTNSTSTVAINSRLLVKEGQVNGKQKIAERSSDADRVLQYF 213
 QY 307 RISHNGVYVDPPLTVL 321
 Db 214 EIRONGRPLDPLSL 228

RESULT 6

Q823Y0 PRELIMINARY; PRT; 250 AA.
 AC Q823Y0;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Possible lipoprotein.
 GN STY3194.
 OS Salmoneilla typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmoneilla.
 CX NCBI_Taxid=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RA MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
 RA Baker S., Baeham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque N., Hien T.T., Holroyd S., Jagels K.,
 RA Kirogh A., Larsen T.S., Leather S., Moul S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmoneilla
 RT enterica serovar Typhi CT18."
 RL Nature 413:848-852(2001).
 DR EMBL; AL672777; CAD02868.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SMO0257; LysM; 1.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 250 AA; 26426 MW; 9A57CF6117B76C4 CRC64;

Query Match 13.9%; Score 232.5; DB 16; Length 250;
 Best Local Similarity 24.4%; Pred. No. 2.7e-09;
 Matches 76; Conservative 49; Mismatches 99; Indels 87; Gaps 9;
 QY 18 LGLIFGVITTCILAGCASKPTYNSTSGSGSHRTSGSGLAIGSQVITDSQGVPRRYOVQ 77
 Db 11 LGLIALLCTGLLGGC-----SSNSGSGTY--SGS-----VYTVKR 44
 QY 78 GDTVSKIAQRYGLNWRREIGHINNLSSTYITGQMLTMSGDLK---VRESISGVT 133
 Db 45 GDTLYRSRATGTSVKELARLNGISPPYITVEGQRIKV-RGSAKSSSTRTKTSNKTATKT 103
 QY 134 AHT-PSVAVQSSRPVQOHPAVOKPFPVAVVKKPPTPVPVQAPAPVPTEA 192
 Db 104 AAVRPSSSVPKSSWPPVQRCWVPANGKVL-----PYST 139
 QY 193 GSSGVMQFRYPVAGATNPVPRFGATVAGSTVNSGMMFSGRGDDLINASAGTVIQADH 252

Db 140 AEGG-----NKGIDINAGRTPVYAAAGKVVYVGN 170
 QY 253 NMDGAS--IVYQHTNGFVSSYIHKDAQVKTGDTVTRTGQRIASMKQPSGALPEFRISR 310
 Db 171 QLRGYGNLIMIKHEDYITAYANDTMYLVNNGQSVAKGQKIAETMGSTDAASVRLHFQIRY 230
 QY 311 NGVYVDPPLTVL 321
 Db 231 RATAIDPLRYL 241

RESULT 7

Q82M83 PRELIMINARY; PRT; 252 AA.
 AC Q82M83;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative metalloendopeptidase.
 GN STM3038.
 OS Salmoneilla typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmoneilla.
 CX NCBI_Taxid=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Lt2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Portnoy L.S., All J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmoneilla enterica serovar Typhimurium
 RT LT2."
 RL Nature 413:852-856(2001).
 DR EMBL; AE008839; RA21913.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SMO0257; LysM; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 252 AA; 26652 MW; 6ADB04D76BAEE61E CRC64;

Query Match 13.9%; Score 232.5; DB 16; Length 252;
 Best Local Similarity 24.4%; Pred. No. 2.8e-09;
 Matches 76; Conservative 49; Mismatches 99; Indels 87; Gaps 9;
 QY 18 LGLIFGVITTCILAGCASKPTYNSTSGSGSHRTSGSGLAIGSQVITDSQGVPRRYOVQ 77
 Db 13 LGLIALLCTGLLGGC-----SSNSGSGTY--SGS-----VYTVKR 46
 QY 78 GDTVSKIAQRYGLNWRREIGHINNLSSTYITGQMLTMSGDLK---VRESISGVT 133
 Db 47 GDTLYRSRATGTSVKELARLNGISPPYITVEGQRIKV-RGSAKSSSTRTKTSNKTATKT 105
 QY 134 AHT-PSVAVQSSRPVQOHPAVOKPFPVAVVKKPPTPVPVQAPAPVPTEA 192
 Db 106 AAVRPSSSVPKSSWPPVQRCWVPANGKVL-----PYST 141
 QY 193 GSSGVMQFRYPVAGATNPVPRFGATVAGSTVNSGMMFSGRGDDLINASAGTVIQADH 252
 Db 142 AEGG-----NKGIDINAGRTPVYAAAGKVVYVGN 172
 QY 253 NMDGAS--IVYQHTNGFVSSYIHKDAQVKTGDTVTRTGQRIASMKQPSGALPEFRISR 310
 Db 173 QLRGYGNLIMIKHEDYITAYANDTMYLVNNGQSVAKGQKIAETMGSTDAASVRLHFQIRY 232
 QY 311 NGVYVDPPLTVL 321
 Db 233 RATAIDPLRYL 243


```

RESULT 8
ID 09L528 PRELIMINARY; PRT; 269 AA.
AC 09L528;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, last sequence update)
DE 01-JUN-2001 (TEMBLrel. 17, last annotation update)
GN NLPD.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KT2440;
RA Ojangu E., Tover A., Kivisaar M.;
RT "Sequence of Pseudomonas putida nlpd gene (complete sequence) and ORF
RT similar to E. coli pcm gene (partial sequence).";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF260132; AAF70311.1; -.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR Pfam: PF01476; LysM; 1.
DR Pfam: PF01551; Peptidase_M37; 1.
DR SMART: SM00257; LysM; 1.
SQ SEQUENCE 269 AA; 2836 MW; F35DCAB481B6B14 CRC64;

Query Match 13.7%; Score 228.5; DB 2; Length 269;
Best Local Similarity 29.2%; Pred. No. 5.6e-09;
Matches 81; Conservative 36; Mismatches 95; Indels 65; Gaps 12;

QY 29 ILAGCASKPTVNTSSGSGSRITSGSGLAIGSQVITTSQGVPR-----YQKQDPTV 81
DB 4 LTVGCSS-----TSN-----SARVVDNNNTVPKPAVTSGQYIVKPDTL 44
QY 82 SKIAQRYGLMREIGHINNINSYTIYTGQWLTLSGDLKVERSSISGVNTAETPPVA 141
DB 45 FSIAPFRGMVYKELANNGOAPYTIIPQAIRSSSS-----SRRTVVSSPS--- 93
QY 142 VQSSRPVQHPAVQKPTPPVVVVKPTPTPPVQDPAPVAPVTEAPPTGSSGVNQFR 201
DB 94 -SSRTVTTRPVGSTATAPASTSKPATSA-----SAPAPVAVPAERAVG-----WT 144
QY 202 YPVGATNPVRRRGTAIVASTVTSNGMPSGRD--GDL--INASAGVITQADHMDG 256
DB 145 WP--ANGVLLGK-----ASNGSLNKGIDIDAGLQDPVPAADGAVVYAGSGLRG 192
QY 257 AS--IVIOHTNGFVSYIHIKDAQVKTGDTVTRTGQRI 291
DB 193 YGELIITIKHSDTVSAVGHNRLLVREGQVKAQSI 229

RESULT 9
ID 09K120 PRELIMINARY; PRT; 230 AA.
AC 09K120;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, last annotation update)
DE Lipoprotein precursor NLPD.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Coxiella group; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21340372; PubMed=11447163;
RA Seshadri R., Samuel J.E.;
RT "Characterization of a Stress-Induced Alternate Sigma Factor, Rpos, of
RT Coxiella burnetii and Its Expression during the Development Cycle.";
RT Infect. Immun. 69:4874-4883(2001).

```

```

DR EMBL: AF244357; AAF73517.1; -.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR Pfam: PF01476; LysM; 1.
DR Pfam: PF01551; Peptidase_M37; 1.
DR SMART: SM00257; LysM; 1.
SQ SEQUENCE 230 AA; 25296 MW; 48312658B04F8533 CRC64;

Query Match 13.6%; Score 227.5; DB 2; Length 230;
Best Local Similarity 28.5%; Pred. No. 5.6e-09;
Matches 72; Conservative 31; Mismatches 83; Indels 67; Gaps 8;

QY 73 YQVKGDPVSKIAQRYGLMREIGHINNINSYTIYTGQWLTLSGDLKVERSSISGVN 132
DB 40 YRVKQGDPTVSIAMAFDLYRALAANRLSPRIEAGQTLQW-----TIPRG-- 88
QY 133 TAHTPSFVAVQSSRPVQHPAVQKPTPPVVVVKPTPTPPVQDPAPVAPVTEAPFAT 192
DB 89 -AHVSGRFA-----AAPRQSPPLQP----- 109
QY 193 GSSGVNQFRYPVGNTPVRRFGATV--AGSTVTSNGMPSGRDGLINASNAGTVIOA 250
DB 110 ----VAHRWPAAG-----RLIGYSVGMAG-----NIGINAGHYGAVRPAASGVVYS 156
QY 251 DHNMDGAS--IVIOHTNGFVSYIHIKDAQVKTGDTVTRTGQRIASMKNOPSGALFEFRI 308
DB 157 GAGIRGYGNLIYKHNTTYISAVAFNRKRVLYKGSRRVRAQKIAEMGRTSGRVMLHFEI 216
QY 309 SRNGVTVDPPLTVL 321
DB 217 RRGQGPVNPRLYL 229

```

```

RESULT 10
ID 09MWV6 PRELIMINARY; PRT; 244 AA.
AC 09MWV6;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, last annotation update)
DE Lipoprotein (Fragment).
GN NLPD.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS358;
RX MEDLINE=20135599; PubMed=10673044;
RA Kotlic M., Degraess G., Venturi V.;
RT "Cloning and characterisation of the rpos gene from plant growth-
RT promoting Pseudomonas putida WCS358: Rpos is not involved in
RT siderophore and homoserine lactone production.";
RL Biochim. Biophys. Acta 1489:413-420(1999).
DR EMBL: Y19122; CAB46190.1; -.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR Pfam: PF01476; LysM; 1.
DR Pfam: PF01551; Peptidase_M37; 1.
DR SMART: SM00257; LysM; 1.
FT NON TER 1
SQ SEQUENCE 244 AA; 25191 MW; 6F9B5BA39CD739AE CRC64;

Query Match 13.6%; Score 226.5; DB 2; Length 244;
Best Local Similarity 28.5%; Pred. No. 7.2e-09;
Matches 82; Conservative 43; Mismatches 96; Indels 67; Gaps 13;

QY 20 LIRGVITTCILACGASKPTVNTSSGSGSRITSGSGLAIGSQVITTSQGVPR----- 72
DB 3 LVIALMGTLTLAGCSSR--SSFS-----ARVVDNRNNAAPKRPYTSQG 43
QY 73 YQVKGDPVSKIAQRYGLMREIGHINNINSYTIYTGQWLTLSGDLKVERSSISGVN 132

```

Db 44 YIVKRGDILFLSLAFPIGMDYKELLAARNGILPAPYTIIRPGQIRPSG-----STGST 94

Qy 133 T-ARTSPVAAVOSSRPVQGHPAVQKFTPPVVVYKKFTPPPPVQQAPAPAPVTEAPFA 191

Db 95 TVSSPS-----SSSTVYLRPVGSGASAPASTTKATYAPFI--PAPVATVPAERA 147

Qy 192 TGSSGVMQFRYPVGATNPVRRFGTATVAGSTVSNQMFMFGSD--GDL--INASNAGT 246

Db 148 VGG-----WTWP--ANGVLGRF-----ASNGSLNKIGDIDAGDQGVPAASGA 190

Qy 247 VIQADHNNDGAS--VIGCHTNGFVSYSYIHKDAQVTKGTPTVRCGLIA 292

Db 191 VVYDGSGLRGVAEIILIKHSIDTVASVAGNRRLLVBEQGQVKAQGSTA 238

```

RESULT 11
093AG4
ID 093AG4 PRELIMINARY; PRT; 217 AA.
AC 093AG4;
DT 01-DEC-2001 (TEMBRLrel. 19, Created)
DT 01-DEC-2001 (TEMBRLrel. 19, Last sequence update)
DT 01-JUN-2002 (TEMBRLrel. 21, Last annotation update)
DE Murelin endopeptidase.
GN MRP
OS Acetobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Azotobacter
CX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UM.
RA Kujat Choy, S.L., Meakins D., Tindale A.E., Page W.J.,
RT "The stationary phase sigma factor (Ipos) of Azotobacter vinelandii."
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF421351; AAL16090.1; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
QQ SEQUENCE 217 AA; 22675 MW; A5AAEB0F64886945 CRC64;

```

Query Match	13.2%;	Score 221;	DB 2;	Length 217;
Best Local Similarity	28.9%;	Pred. No. 1.5e-08;		
Matches	73;	Conservative	92;	Indels 58; Gaps 8
		Mismatches	92;	

```

QY      87  YUGLWLRRELGNNINSSSTIYTGQWLLTMSGDVKVRSISISSGVTANTHPSPVAQSSR 146
Db      2  RYGMWRRLAANNGIAAPVHHPGQKIQI----- 30

QY      147  PRVOONHNVQKTRPRVVUVKKRPRPRVVUQAPVA-----RPVEAPRATGS 194
Db      31  GPVRQ-DAASTARBSVRATAPRYATSGAAVAATPVKPSDCKQVPPATATPSA-- 86

QY      195  SCVMQFRPRVQ---ATNPVRPRFGATYAGSVTNSGMIFFSGRDEDLINASNAQYIQA 250
Db      87  OPVARSBSGAMPREGLVGRFS--NGS--LNGKIDIGDGLQGPVLAADSGSVYA 1411

QY      251  DHNMDSAG--IYIQTNGFVSSTIHKDAQVKTGDTVTRGORTASKMOPSGALFEERI 3089
Db      142  GGGELGYGELLIIKISDVTYVSAIGHNRLLVREGQGVKAGQVIAEMSGSTGIDVKLHEI 201

QY      309  SRNGVYVDPPLTVL 321
Db      202  RRGKRPVDPLOYL 214

```

RESULT 12		
ID	PRELIMINARY;	PRT; 268 AA.
08Y038		
08Y038		
08Y038;		
01-MAR-2002	(TREMblrel. 20, Created)	
DT		
01-MAR-2002	(TREMblrel. 20, Last sequence update)	

01 JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Probable lipoprotein NLPD/LPPB homolog.
NM RSCG1206 OR R802686.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia
NCBI_TaxID=305;
[1]
RN NCBI_TaxID=1299;
RM [1]

SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RA MEDLINE=21681879; PubMed=11822852;
RA Salanoudat M., Genin S., Attiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brotilier P., Camus J.C., Catecholice L.,
RA Chandelier M., Choisme N., Claudel-Renaud C., Guinac S., Demange N.,
RA Gaspin C., Lavie M., Molisan A., Robert C., Sautin W., Schlex T.,
RA Siyler P., Thebault P., Whalen M., Winkler P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequencing of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
DR EMBL; AL646063; CAD14908.1; -
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002816; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 1.
DR PROSITE; PS00782; FLIIB; UNKNOWN_1.
SQ Complete proteome.
Sequence 268 AA; 27918 MW; EC0386D8E8367CE CRC64;

Query Match 13.2%; Score 220; DB 16; Length 268;
Best local similarity 24.9%; Pred. NO. 2.4e-08;
Matches 78; Conservative 45; Mismatches 133; Indels 59; Gaps 8

14 PIKSLGLIFGYITCTILAGCAK---PTVSTSGSGSHRISGSGALIGSOVLINDSGV 69
Db 6 PARAGRLLAVNVSAALLAACSSGNQAPVDRIIRAGS-----APAFLERPP 53
QY 70 PNRIYQKGGDTVSKIAQRYYGLNMRLEIGHINNINSYTYTQOMLTWMSGDLKRRERSISS 129
Db 54 PGYRKRGKGDPLYSTALNNGGAPRPDLVTANNIPENNQLVEDOLRV-----VPNNADLS 107
QY 130 GVNTAHTPSVAVOSSRPDPQHAPAKFPPEPVVVVKFPTPEPVYQOPAPVAPPYTEAR 189
Db 108 ATGAVALP-----VRPP-----NTTQPLDAAPATFPVVSSEA 141
QY 190 FATGSSGVMDQERYPGVATNPVRRFGTAIVAGSTVTSNGMWFSGEDGLINASNAVYIQ 249
Db 142 SAGATDGAIALAMP--AHGOVIYGR-----DDKANGLIDIGCKRGCAVLAADDGKYIH 192
QY 250 ADHMMND-GASIVIOHTNGFVSYSYHIKDQAVYKTGPDVTARTGORIASMKNQPSGALFFERI 308
Db 193 VGPLRGYNLVLIKXNDLFETLAGNDRKVLVBOSTVKKGIKIAMSGSTDA DRVYLAFHEV 252
QY 309 SRNGVYVDPLTVL 321
Db 253 RRNGKVPVDMRF 265

RESULT 13
O9RS36 PRELIMINARY; PRT; 454 AA.
AC O9RS36;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DB Cell wall glycy-L-glycine endopeptidase, putative.
GN DR2291.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococciales;
OC Deinococcaceae; Deinococcus
OX NCBI_TaxID=1299;
RM [1]

Complete proteome.
 KM SEQUENCE 311 AA; 33507 MW; FC70D9EF1396451E CRC64;
 Query Match 12.5%; Score 208.5; DB 16; Length 311;
 Best Local Similarity 25.8%; Pred. No. 2e-07;
 Matches 85; Conservative 48; Mismatches 139; Indels 59; Gaps 12;

QY 17 RLGLIFGITTICLAGCAKSKPTNSTSGSGSHRTSGGLAIGSVITDSQGVNRYQK 76
 DB 7 RLGLL---LFCSLFLGCTA-PIPAVPSGLKDYKYNKEVSGYRGST-----YKVK 51
 QY 77 QGDIVSKIAQRYGALNWEIRGHINNNINSYITYGWMLTW-----SGDLK-- 122
 DB 52 KGDILYFAYLITDKQVNDLISYNDLAPYTIHFGOKITLMLPNYTPPAYGCTGATVAV 111
 QY 123 ---RERSISSGVNTAHTPSPVAVOSSRPPVQ-----QHPAVQKPTPPVVKKPTPTP 173
 DB 112 ASSTASAVAKAATATVAQTVAQTSKSNNTKVNQNSTNSQNLTKQDVKTDQTKKEYEP 171
 QY 174 VVOQPAFV---APVTEAPFATSGSGVMQFRYPVGATNPVRRFGATVAGSTVTSNGM 229
 DB 172 VQKONVNVNVAKAKPSDEK-----IAKWLMP--TKGRVYKNFS---AGDQ-GNKGI 216
 QY 230 WFSGRDGLINASNAGTVIOAHNMMDGAS--IVIOHTNGFVSSYTHIKDAQVKTGDTVRT 287
 DB 217 DIAGORGAVVAATDAGTVVSGNALRGVGNLIHKGHEHYLSAVAHNDQLAKEGQTVQA 276
 QY 288 GORIASMKNQPSGAALFEFRISRNQVYDP 317
 DB 277 GOKIATMGSGSTNSVRLHFEIRYQOKSYNP 306

RESULT 16
 ID 08YGS9 PRELIMINARY; PRT; 427 AA.
 AC 08YGS9
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Lipoprotein NLPD.
 GN BMEI1079.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 RX NCBI_TaxID=29459;
 RT [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DeIvecchio V.G., Kaparatel V., Redkar R.J., Patra G., Mujar C., Los T.,
 Ivanova N., Andersen I., Bhattacharyya A., Lykidis A., Resnik G.,
 Jalonoski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 Harelkorn R., Kyriakis N., Overbeek R.;
 RA "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 RL EMBL; AB009547; AL152260.1; --
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 KW Complete proteome.
 SQ SEQUENCE 427 AA; 43604 MW; 48912D48846EBD7 CRC64;

Query Match 12.3%; Score 205; DB 16; Length 427;
 Best Local Similarity 26.2%; Pred. No. 5.4e-07;
 Matches 79; Conservative 45; Mismatches 132; Indels 46; Gaps 12;

QY 30 LAGCASKPTNSTSGSGSHRTSGGLAIGSVITDSQGVNRYQKQDVTYSKIAQRYG 89
 DB 162 IAGVQAPAVNAGKSSPTNMAAGCAIA-----TPSVGAYVVKSGDSIFSIKKN 214

QY 90 LNMREIGHINNINSYITYGWMLTWSGDLKVERISSGVNT---AHTSPV-AVOSS 145
 DB 215 VPIVDQLKRAAGLTSG-AIRVQOGLVI-----PSAAGAPQVAIFPOANPKAA 264
 QY 146 RPPVQGHVAVQKPTPPVNVVKKPTPTPPVVOQAPAPVPTAPFATGSSGVMQFRYPVG 205
 DB 265 SAPANGAHPVKTPTPPQASNK-----VIEDAE-----KQQAAPSSTISQMWPFY- 310
 QY 206 ATNPVRRFGATVAGSTVTSNGMFGSRGDDLINASNAGTVIOAHNM--GASITQH 263
 DB 311 -RGLHSFGQRE---GTSVSDGIDIVPEETPVKALENGVIYAGGGLKEFGQTVLIRH 366
 QY 264 TNGRVSSYTHIKDAQVKTGDTVRTGQRIASMKNQPSGAA--LFEFRISRNQVYDP 320
 DB 367 DNGLVTVYGHNSQITVQVQKVRKGEEL--KSGMSGNASPKLHFEVRKNSAPVNEFTKY 424
 QY 321 LK 322
 DB 425 LE 426

RESULT 17
 ID 09UTP1 PRELIMINARY; PRT; 415 AA.
 AC 09UTP1
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Putative membrane peptidase.
 GN NMA1692.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 RX NCBI_TaxID=65699;
 RT [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=2022556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Kottrelli G., Basham D., Brown D., Chillingworth T.,
 Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S.,
 Jørgensen K., Leather S., Moule S., Mungall K., Quail M.A.,
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 Whitehead S., Spratt B.G., Barrall B.G.;
 RA "Complete DNA sequence of a serogroup A strain of Neisseria
 RT meningitidis Z2491."
 RT Nature 404:502-506(2000).
 RL EMBL; AL162756; CAB84920.1; --
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 2.
 KW Complete proteome.
 SQ SEQUENCE 415 AA; 42883 MW; B39EBA52C7CD78F7 CRC64;

Query Match 12.3%; Score 203; DB 16; Length 415;
 Best Local Similarity 27.3%; Pred. No. 7.2e-07;
 Matches 71; Conservative 42; Mismatches 97; Indels 50; Gaps 11;

QY 73 YQVQGDVTYSKIAQRYGALNWEIRGHINNNINSYITYGWMLTWSGDLKVERISSGVN 132
 DB 185 HTVIRGDIYVNIKSRHYHISODDFPAMGMNTN-RLTSLGQ-----IVKVPAGYAA-- 233
 QY 133 TAHTPSPVAVOSSRPPVQGHVAVQKPTPPVNVVKKPTPTPPVVOQAPVAD----- 183
 DB 234 ---PKAAAVK-SRPAPV--AAVQTPV-----KPAQAPVQSAFPQAAAPAKNAKVA 279
 QY 184 PVTAPFPA---TSSSGVMQFRYPVGATNPVRRFGATVAGSTVTSNGMFGSRDGLI 239
 DB 280 PAPQSPAPSPGCTRSVGGIWMQRP--TQGKVVAQFGG-----NNKGVADIAGNAGQPV 329
 QY 240 NASNAGTVIOAHNMMDGAS--IVIOHTNGFVSSYTHIKDAQVKTGDTVRTGQRIASMKNQ 297

Db 330 LAADGCVVYVAGSGLRVGNLVITIOHNSFLTAIGHNOKLVGBGQGVKGGQVLAIMGNT 389
 Qy 298 PSGALLFEFRISRNQVYVDP 317
 Db 390 EASRTQLHFEVRONGKFPVNP 409

RESULT 18

Q9X7Z3 PRELIMINARY; PRT; 379 AA.
 AC Q9X7Z3; 08X7Z3;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Lipoprotein.
 GN NLPD OR Z4050 OR EGS3596.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxId=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grobeck J., Davis N.W., Lim A., Dymally E.T., Potamousis K.,
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-g., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
 Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12,"
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AE005502; AAG57849.1; -
 DR EMBL; AP002562; BAB37019.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM.1.
 DR Pfam; PF01551; Peptidase_M37.1.
 DR SMART; SM00257; LysM.1.
 KW Complete proteome.
 SQ SEQUENCE 379 AA; 40172 MW; F664459AID90720 CRC64;

Query Match 12.1%; Score 202.5; DB 16; Length 379;
 Best Local Similarity 26.1%; Pred. No. 7e-07;
 Matches 72; Conservative 42; Mismatches 113; Indels 49; Gaps 9;

Qy 73 YQVQGDPTVSKIAQRYGLNMRREIGHNINNSYTIYTGQWLTMSGDLKYRERSISSGVN 132
 Db 123 YTVKQKDTLFTYIMTIGNDFRLAQRNINQAPYALNVGQ--TLQVG-----N 167
 Qy 133 TAHTPSPVAVQSRPPVQGHPAVQKP--TPPVVVVKKPT-----P 170
 Db 168 ASGTPTTGGNATIQADAAEGGVVAKPQNSTVAVASOPTTYSSESGEGANMLPNK 227
 Qy 171 TPPVQOPAPVAPVTEAPFATG---SSGVQFRYPGATNPVVRFGATVAGSTVTSN 227
 Db 228 TATVTAPVTPVPASTTEPTVSTSTSTPSTRWP--TEGKVIETFG-ASEGG---NK 280
 Qy 228 GMMFSGRGGDLINASAGTYTQADHNMDS--IVIQHTNGFVSSYTHIKDAQVKTGDTV 285
 Db 281 GIDIASGSGQAIATADGRVYVAGNALRGYGNLIIIKHNDVYLSAVAHNDTMLVREQEV 340

Qy 286 RTGQRASMKQPSGALFEFRISRNQVYVDP 321
 Db 341 KAGQKATMGSTGTSRHLFEIRYKKSVPRLYL 376

RESULT 19

Q9JYP9 PRELIMINARY; PRT; 415 AA.
 AC Q9JYP9; 09JYP9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Lipoprotein NlpD, putative.
 GN NMB1483.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B.
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Ciecho A., Parksey D.S., Blair E., Citrone H., Clark E.B.,
 Cotton M.D., Uetreback T.R., Khouri H., Qin H., Yamatchevan J.,
 Gill J.O., Scarlato V., Masiagnani V., Plaza M., Grandi G., Sun L.,
 Smith H.O., Fraser C.M., Moxon E.R., Rappaport J., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58,"
 RL Science 287:1809-1815(2000).
 DR EMBL; AE002497; AAP41839.1; -
 DR TIGR; NMB1483; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM.2.
 DR Pfam; PF01551; Peptidase_M37.1.
 DR SMART; SM00257; LysM.2.
 KW Complete proteome.
 SQ SEQUENCE 415 AA; 42875 MW; 7704BF7F0DF54DF5 CRC64;

Query Match 11.9%; Score 198; DB 16; Length 415;
 Best Local Similarity 27.2%; Pred. No. 1.7e-06;
 Matches 72; Conservative 38; Mismatches 95; Indels 60; Gaps 12;

Qy 73 YQVQGDPTVSKIAQRYGLNMRREIGHNINNSYTIYTGQWLTMSGDLKYRERSISSGVN 130
 Db 185 HTIVRGDTVYNISKRY-----HISQDD-----FRAMNG---MTDNNLSIGOI 223
 Qy 131 --VNTAHTPSP--VAVQSRPPVQGHPAVQKPPTPVVVVKKPTPTPVVQOPAPVAP--- 183
 Db 224 VKVPAQGYAPKPAVAESRPAPV--AAVQTPV-----KPAQPPVQSAFQPAAPAAEN 274
 Qy 184 -----VTEAPFA-----TGSSGVQFRYPGATNPVVRFGATVAGSTVTSNGMMSGR 234
 Db 275 KAVPAPAPQSPAPSPGCTRSVSGITVWRP--TGCKVADBGG-----NNKGVDIAGN 324
 Qy 235 DGDILINASAGTYTQADHNMDS--IVIQHTNGFVSSYTHIKDAQVKTGDTVTRGQRIA 292
 Db 325 AGQPVLAADGKVVYVAGSGLRVGNLVITIOHNSFLTAIGHNOKLVGBGQGVKGGQVLA 364
 Qy 293 SMKNQPSGALLFEFRISRNQVYVDP 317
 Db 385 LMGNITDASRTQLHFEVRONGKFPVNP 409

RESULT 20

Q98LD1 PRELIMINARY; PRT; 515 AA.
 ID Q98LD1
 AC Q98LD1;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)

DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE Lipoprotein.
 GN MLI1077.
 OS Rhizobium loti (Mesorhizobium loti).
 OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OC NCB1_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneke T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tadota S.
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP002996; BAB48532.1; -.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 2.
 KW Complete proteome.
 SQ SEQUENCE 515 AA; 52902 MW; 0837B06F79AAFBDD CRC64;
 Query Match 11.8%; Score 196.5; DB 16; Length 515;
 Best Local Similarity 25.7%; Pred. No. 2.8e-06;
 Matches 69; Conservative 38; Mismatches 109; Indels 53; Gaps 10;
 Oy ITDSGVNRYOVKQDTVSKIAORYGLNWEIGHINLNSYTYTGOWLTWMSGDLKV 122
 Db 281 VAKATGAGTGTGSGDITMSIARTKGTGVAALAKANMKOG-LIKIQTLKVPKG--- 335
 Oy 123 RERSISGVNHTSPVAVOSSRPVVOCHPAVQKPTPPVVKPTTPVVOQ---PA 179
 Db 336 -----GTATVASSKPAKVDPVTATTOPAKTTPSETLSTYPPKDAKVIQAEDE 388
 Oy 180 PVAPPTVAPPAATGSSGVMOFRYPVAGATNPVRRFGATVAGSTVTSNGMFGSGDG-DL 238
 Db 389 AVAPDAT-----GIGKMKWPV--RGVVISGFG-----SGDGVDI 421
 Oy 239 -----INASNAGTVIOADHNMD--GASIVIQHTNGFVSYTHIKDAQVKTGDTVRTGQ 289
 Db 422 AVPGSTPTKAENGVIYAGDGLKEFGNTVVRHENGIVTYGHAASIEVORGOKVRGQ 481
 Oy 290 RIA-SMKNOPSGAALFEFRISRNGVYVDP 317
 Db 482 EIALSGMSGTTDSPGLHFEVVKNSAPVDP 510
 RESULT 21
 ID 09A6T7 PRELIMINARY; PRT; 609 AA.
 AC 09A6T7;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Peptidase; M23/M37 family.
 GN CCI996.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OC NCB1_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A. / CB15;
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwim M.L., Halt D.H.,
 RA Kolonay J.F., Smt J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Uterback T., Tran K., Wolf A., Vamathevan J., Emolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
 RT "Complete genome sequence of Caulobacter crescentus."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AB005873; AAK23971.1; -.
 DR HSSP; P23931; IEQG.
 DR TIGR; CC1996; -.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 4.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 4.
 KW Complete proteome.
 SQ SEQUENCE 609 AA; 63071 MW; 6C6EPA0C426BB34 CRC64;
 Query Match 11.7%; Score 196; DB 16; Length 609;
 Best Local Similarity 23.6%; Pred. No. 3.8e-06;
 Matches 66; Conservative 54; Mismatches 110; Indels 50; Gaps 12;
 Oy 73 YOVKQDVTASKIAORYGLNWEIGHINLNSYTYTGOWLTWMSGDLKVRERSISGVN 132
 Db 347 YSVQGTDLGRIARFNVSVKALAEENLRATSLKKGQKIALPDG---FRDK---GPIR 400
 Oy 133 TAHTPSVAVOSSRPVVOCHPAVQKPTPPVVKPTTPVVKPTTPVVKPTTPVVKPTTPV 188
 Db 401 TTTTTRP-----ATPANTYARVDSAAASTPSPVPTPSGAAPSPAPVAAQPTTP 455
 Oy 189 PFATG-----SSGVMOFRYPVAGATNPVRRFGATVAGSTVTSNGMFGS 232
 Db 456 P-SGGRTIIEAPPAATEAIIASGKGKAWPL--RGIISFG---YKGGQRNDGLNIR 509
 Oy 233 GRODDLINASNAGVIOADHNMD--GASIVIQHTNGFVSYTHIKDAQVKTGDTVRTGQ 290
 Db 510 APQGTPLVSSADGEIAIANGOVPTFGULVLVHADGVTVIAHLSTNNVKKRQVKGQEQ 569
 Oy 291 IASW-----KNQPSGAALFEFR-----ISRNGVYVDPITVL 321
 Db 570 LGIVGATGVNEPO--LHFERIYAPTVKDYAKPVPALVL 607
 RESULT 22
 ID 08UEO5 PRELIMINARY; PRT; 562 AA.
 AC 08UEO5;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE Lipoprotein.
 GN ATU1700 OR AGR C.3124.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OC NCB1_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Sebubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D., St.
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-o., McClelland E., Palmieri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Neeter E.W.,
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58."
 RL Science 294:2317-2323(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194; Miller N., Blanchard M.,
 RA Goodner B., Hinkle G., Gattung S., Cao Y., Akenazi M., Halling C., Mullin L.,
 RA Ourullo B., Goldman B.S., Garton J., Tatchouk O., Epp A., Liu P.,
 RA Houmel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu P.,
 RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Garton J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL; AE009126; AAU42700.1; ALT_INIT.
 DR EMBL; AE008091; AAK87473.1; .
 KW Complete proteome.
 SQ SEQUENCE 562 AA; 57218 MW; B6479E31F579A84 CRC64;

Query Match 11.3%; Score 188.5; DB 16; Length 562;
 Best Local Similarity 25.1%; Pred. No. 1.2e-05;
 Matches 81; Conservative 64; Mismatches 131; Indels 47; Gaps 15;

QY 5 IAINSONOKPIKRLGIFGVITTCILAGCAKPTYNSTSGSGSHRTSGGGLA-IGSQVI 63
 DB 272 IALNRNDGP-----PLKRAPDGNVAVLPQAAARDKLSSGAKLTTPGCKPL 319
 QY 64 TDSQGVNRYQVKQGDVTSKIAQRYGLMREIGHINNLSYTYTGQWLTLMGDLKVR 123
 DB 320 PPSGG-----YKVGQDLSIAKIRANGVSVALKANGI-SNESIRVGOTLTM----- 366
 QY 124 ERSISGVTNHTSPVAVQSSRPVQHPVQKPTP---PVVVKKPTPPPVQOPAP 180
 DB 367 PGASTDITKITAASVAKKA--AAAKPVE--TASIKPEPKAPAAATPAATPATASVD 422
 QY 181 VAPVTEAPFATSGSGVQGFYRV-GATNPVRRFGTATVAGSVTNSGMFSGRDGLI 239
 DB 423 IEKKADMASTAPSTGIGKTRMPVPGA--VINFGD-NVGS--KNDGINISVDEGPI 476
 QY 240 NASNAGTVIQADHNM--GASIVIOHTNGFVSYIHIDKADVKTGTVRTQRIASMKQ 297
 DB 477 KAANGVVIVYAGNGLKQLGNTVLVRHDDGKVTGVGNANLVDVGQVQVROQTAT--SG 534
 QY 298 PSCAA---LFEFRISRNGVYVDP 317
 DB 535 MTGSAAKRPQVHFVVRKDATPVNP 557

RESULT 23
 008251
 ID 008251 PRELIMINARY; PRT; 512 AA.
 AC 008251;
 DT 01-JUL-1997 (TREMblrel. 04, Created)
 DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Lipoprotein.
 GN LPPB.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=97449848; PubMed=9304864;
 RA Streit W.R., Phillips D.A.;
 RT "A biotin-regulated locus, bios, in a possible survival operon of
 RT Rhizobium meliloti.";
 RL Mol. Plant Microbe Interact. 10:933-937(1997).
 DR EMBL; U81296; AAB88077.1; .
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 SQ SEQUENCE 512 AA; 52431 MW; 0A870A76BCE7D8B CRC64;

Query Match 10.4%; Score 174.5; DB 2; Length 512;
 Best Local Similarity 23.5%; Pred. No. 0.00011;
 Matches 95; Conservative 58; Mismatches 128; Indels 123; Gaps 19;

QY 24 VITTCIAGCASKPTYNSTSGSGSHRTSGGGLAIGSQVITDSQGVNRYQVKQDVTYSK 83
 DB 127 VQTASLSQAAALPKAPPLAGAKD-MSGKQGW-----ASNAP-IMVRGDDFTV 176
 QY 84 IACRYGLMREIGHINNLSYTYTGQWLT-----SGVNTAHTSP----- 113
 DB 177 LARFVGEKEITKANLKSASQVEPQRLVPTFTGASAKAASGSIAADVEGKKRP 236
 QY 114 -----TLMGDLKVRERSIS-----SGVNTAHTSP----- 139
 DB 237 SPLPTDQREVAAILPQGSQSRKESRSDVAAGKLSAGEGGNGAITYKPEDSLNRITAKA 296
 QY 140 -----VAQSSRP--PVQHP--AVQKPTPPVV-----VKK-----PTPTP 172
 DB 297 NGVFGRCPEAGKRAFDRSHPRAKTQHPFLGENTGTDAVTVASAKKNEAQAASFTQG 356
 QY 173 PVQGPAPVA-PVYTEAPFATG-----SGVQGFYRV-GATNPVRRFGTATVAGST 223
 DB 357 KLITETKAPAKESVEVAIRSDGNEDELKSTGIGKYPVARGA--VVAAYG-ANVGN- 411
 QY 224 VTSNGMFGSGRDGLINASNAGTVIQADHNM--GASIVIOHTNGFVSYIHIDKAOYKT 281
 DB 412 -RNDGINISVDEGPIIAANGVIVYSGSSIKELGNVLVRHDDGTVTVGNAAELKVR 470
 QY 282 GDTVRTQRIASMKQPSGAAL---FEFRISRNGVYVDPPLTVLK 322
 DB 471 GQKVRGQTLAS--SGMTGRATRPQVHFVVRKDATPVNPATYLE 512

RESULT 24
 0926D2
 ID 0926D2 PRELIMINARY; PRT; 512 AA.
 AC 0926D2;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Lipoprotein precursor.
 GN LPPB OR R01536 OR SMC02060.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Bactut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaire V., Massey D.,
 RA Pohl T., Portetelie D., Puehler A., Purnelle B., Rampeguy U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591787; CAC46115.1; .
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR Pfam; PF01551; Peptidase_M37; 1.
 KW Complete proteome.
 SQ SEQUENCE 512 AA; 52003 MW; 6BBADDD3290125D0 CRC64;

Query Match 10.4%; Score 174; DB 16; Length 512;
 Best Local Similarity 24.0%; Pred. No. 0.00012;
 Matches 79; Conservative 60; Mismatches 122; Indels 68; Gaps 12;

QY 37 PYNSTSGSGSHRTSG-----GGLAIGSQVITDSQ----- 67

Db 209 PTFGTASAKAAAGSADVEGKKRPSPLPTDREVALIPGOSQREKSESRDVAAG 268
Qy 68 -----GYPNRQVQGPVSKIAQRYGLNMEIGHINNINSYTYTGWLTMSGD 119
Db 269 KLASAGGGGNGATVTPGDLNRIKANGSVAAKQANGL-STEARIRQCKINTI9A- 326
Qy 120 LKATRENSISSGVTATHTPPPAVQSSRPVQGHAVQKPTPPVVVKKPTPTPVQOPA 179
Db 327 -----SAKTPADVAVTASVAKKNEQAASTEQKLTG--TKAPAKESVS 371
Qy 180 PVAPVTEAPATGSSGVQGFRRPV-GATNPVRRFGTAVAGSTVTSNGMFGSDGDL 238
Db 372 EVAIRSDGNEDLPKSTIGIKRWVRGA--VVAAYG-ANVDGN--RNDGINISVPGTP 425
Qy 239 INASNACTVIQADHNM--GASIVIOHTNGFVSYIHKDAQVKTGDTVRTGORIASMKN 296
Db 426 IKAENGVVYISGSLKELGNAVLVRHDDGTIVYGNAAELKVGQKVGCRQTLAS--S 483
Qy 297 OPSGAL--FEFRISRNGVVDPPLTVLK 322
Db 484 GMTGRATRPVHFVRKNATPVNPATYLE 512

RESULT 25

08RNB5 PRELIMINARY; PRT; 392 AA.
AC 08RNB5
DT 01-JUN-2002 (TREMREL. 21. Created)
DT 01-JUN-2002 (TREMREL. 21. Last sequence update)
DT 01-JUN-2002 (TREMREL. 21. Last annotation update)
DE NLPD.
GN NLPD.
OS Bartonella henselae (Rochalimaea henselae).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Neisseriaceae; Bartonella.
OX NCBI_TaxID=38323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOUSTON-1.
RA Flakus W., Padmalayam I., Baumstark B.;
RT "Molecular cloning, sequencing, and characterization of the NLPD/lppB
homolog in Bartonella henselae."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF484425; AAL89761.1; -
SQ SEQUENCE 392 AA; 42635 MW; CF71E6B91FDA206C CRC64;

Query Match 10.3%; Score 172.5; DB 2; Length 392;
Best Local Similarity 24.0%; Pred. No. 0.00011;
Matches 79; Conservative 49; Mismatches 130; Indels 71; Gaps 12;

Qy 35 SKPTVNSTSGSGSHRTSGGGLAIG-----SOVITDSQGVNRYOVKQDPTVSKIAQRYGL 90
Db 93 NSPPDGGTSSPNS-RINGTTPRNIGTLSRSQWRNDPLFRQNSYIVQTGDTLISARQGV 151
Qy 91 VWRREIGHINNINSYTYTGWLTMSGD-----KXER 125
Db 152 SVEALIKVNGIRSN-SITIGVLMIPSGRTAETSVNRNDMGASQSLSQSASISHK 210
Qy 126 SISGCVNTAHTPPPAVQSSRPVQGHAVQ-----KPTPPVVVKKPTPTPVQOP 178
Db 211 KYSSSTEKAPITPKPEA-QINNSNGQSSAQMSLNTEGVLDPVKNKONGVTPQ----- 264
Qy 179 APVAPVTEAPATGSSGVQGFRRPVGATNPVRRFGTAVAGSTVTSNGMFGSDGDL 238
Db 265 -----ATGISKRWVR--RGRLLSFGGR--GTTMRGIDIAVPESS 304
Qy 239 INASNACTVIQADHNM--GASIVIOHTNGFVSYIHKDAQVKTGDTVRTGORIASMKN 296
Db 305 VKAENGVVYISGSLKELGNAVLVRHDDGTIVYGNAAELKVGQKVGCRQTLAS--KS 362
Qy 297 OPSG--AALFEFRISRNGVVDPPLTVLK 322
Db 363 GVSQGVKTPRVYFEVRENSLPVDPIKYLE 391

RESULT 26

09X7M8 PRELIMINARY; PRT; 312 AA.
AC 09X7M8
DT 01-NOV-1999 (TREMREL. 12. Created)
DT 01-NOV-1999 (TREMREL. 12. Last sequence update)
DT 01-JUN-2002 (TREMREL. 21. Last annotation update)
DE Putative secreted peptidase.
GN SC06773 OR SC6A5.22.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D.; Parkhill J.; Barrell B.G.; Rajandream M.A.;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinasli H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145.
RA Bentley S.D., Chater K.F., Cerdano-Farraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra J., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Ruter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL: AL049485; CAB39706.1; -
DR InterPro: IPR002482; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR Pfam: PF01476; LysM; 1.
DR Pfam: PF01551; Peptidase_M37; 1.
DR PRINTS: PR01574; TUBBYPROTEIN.
DR SMART: SMO0257; LysM; 1.
SQ SEQUENCE 312 AA; 31633 MW; 2D489B9B7C448206 CRC64;

Query Match 10.3%; Score 171.5; DB 16; Length 312;
Best Local Similarity 23.3%; Pred. No. 9.5e-05;
Matches 79; Conservative 43; Mismatches 118; Indels 99; Gaps 15;

Qy 46 GSHR-----TSGSGGLAIGSOVITDS-----QGP----- 70
Db 5 GKRRRTAMRLRTITVAAGTGAALALPLMGAAAANAAPASVSEQAVQSPAAKAKAA 64
Qy 71 -----NKYOVQGDPTVSKIAQRYGLNMEIGHINNINSYTYTGWLTMSGDLYRE 124
Db 65 EKNSDSRTTYVSSDYSLADEQDV-----GKWKUAYDN-----R 102
Qy 125 RSISGCVNTAHTPPPAV--QSSRPVQGHAVQKPTPPVVVKKPTPTPVQOPAPVA 182
Db 103 EAVSDPSLHHPGKLSTIDGAAPKPSAPSSAQSKFA-----OKSAEKPAEKTAAPPA 155


```

QY 183 PVTAPATGSSGWMQFRYPVGTATNPVRRFGATATVAGSTV-----TSGMWSG----- 233
DB 156 QKSTDAKASSSDPGSQ---STGTTSGY-----TSPVAGGTGVPYHQSGSMSSGVTHTG 207
QY 234 -----RUGDLINASNACTVIOAD-HNMDSASIVQHTNGFVSSYTHIKDAQVKTGDTVRT 287
DB 208 TDFVVPFTGTSLKAGAGTIVSAGGAYGQVTVQLADGHYAQAHLSSLSVAGOSVTA 267
QY 288 GORI--ASMKNQPSGAL-FFRISRN-GVVDPLTVLK 322
DB 268 GQVGLSGATGATGVTGPHLHFIRITPDYGSIDPLAFLR 306

RESULT 27
Q8R6U3 PRELIMINARY; PRT; 449 AA.
AC Q8R6U3;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE Membrane proteins related to metalloendopeptidases.
GN NUPD6 OR TTE2691.
OC Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxId=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; Pubmed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RU Genome Res. 12:689-700(2002).
DR EMBL; AE013208; AAM25810.1; -.
SQ COMPLETE PROTEOME.
KW SEQUENCE 449 AA; 50731 MW; EA7A06CB6CD151F0 CRC64;

Query Match 10.2%; Score 171; DB 16; Length 449;
Best Local Similarity 27.6%; Pred. No. 0.00017;
Matches 74; Conservative 42; Mismatches 120; Indels 32; Gaps 12;

QY 73 YQVQKQDPTSKIAQRYGLNMBREIGHN-----NLNSYTYITGQWLTLMGSDLKVRRRSI 127
DB 193 YEKVENDSLMSIAREHMYIQDILKLPGLTENLKPQIILYLSKEVPLV---VTEKEY 249
QY 128 S-----SGVNTAHT-PSPFAVQSSRPVQOHPAVQKPTPPVYVVKKPTTPPVVQOPAPVA 182
DB 250 TYKBEIPEFTKFTKDDTLFVNSQKVLVEGKGL-KEIKAVIVISHNGVEVRDIEKERVLK 308
QY 183 PVTAPATGSSGWM-----QFRYPVGTATNPVRRFGATATVAGSTVTSNGMWSGSDG 237
DB 309 EPISKI-VAAGSRVSVYATGYSPARGT--ITSRGPWGGFHT---GVDIARYS 361
QY 238 LINAASNACTVIOADHNMND-GASIVQHTNGFVSSYTHIKDAQVKTGDTVRTGORIASMK 296
DB 362 PIASDGGIVIFKMGESGYGLVKIDHNGVTVYTHGASKLAVKVGDKVEKQKIALV-- 419
QY 297 QPSGAL--FFRISRNQVVDPLTVL 321
DB 420 GATGHATGPHVHFVRKNGVPIDMRFL 447

RESULT 28
Q8X3L6 PRELIMINARY; PRT; 223 AA.
AC Q8X3L6;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Putative lipoprotein.

```

```

GN 24203.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; Pubmed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamocitis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RU Nature 409:529-533(2001).
DR EMBL; AE005516; AAG57994.1; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM_1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 1.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 223 AA; 23472 MW; 774D23D63C9C81C3 CRC64;

Query Match 10.1%; Score 169; DB 16; Length 223;
Best Local Similarity 24.0%; Pred. No. 9.3e-05;
Matches 63; Conservative 35; Mismatches 79; Indels 86; Gaps 8;

QY 29 ILACGSKPTYNSTSGSGSHRTSGGGLAIGSVITDSQGVPNRYQVQKQDPTSKIAQRY 88
DB 9 ILACG-----SGSKSSDTGYSG-----SVYTVKRGDTLYRISRTT 44
QY 89 GLAMREIGHNINLNSSTTYITGQWLTLMGSDLK---VRRSISSGVNTAH-TPSPVAVQ 143
DB 45 GTSVKELARLNGISPPYITVEGQKTLK--CGAKSSSTRKSTAKSTYKTASVTPSSAVPK 102
QY 144 SSRPVQGHAVQKPTPPVYVVKKPTTPPVVQOPAPVAPVTEAPATSSGWMQFRYP 203
DB 103 SSMPVQRCWMLPPTTGKVM-----PSTADGG----- 131
QY 204 VGATNPVRRFGATATVAGSTVTSNGMWSGSDGLINASNACTVIOADHNMDSG--IVI 261
DB 132 -----NKGIDISAPRGTPPIYAAGAGVYVYVNGDLRGNGLIMI 169
QY 262 QHTNGFVSSYTHIKDAQVKTGDT 284
DB 170 KHSEDYITAYAHNDTMLVNNNGS 192

RESULT 29
O83190 PRELIMINARY; PRT; 371 AA.
ID O83190;
AC O83190;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Hypothetical protein TP0155.
GN TP0155.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxId=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; Pubmed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Kechum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utecherback T.,
RA McDonald L., Atliach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

```

RA Venter J.C.;
 RT "Complete genome sequence of *Treponema pallidum*, the syphilis
 agent."
 RL Scienc 281:375-388(1998).
 DR EMBL: AF001200; AAC65145.1; -
 DR TIGR: TP0155; -
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR002886; Peptidase_M37.
 DR Pfam: PF01476; LysM; 2.
 DR Pfam: PF01551; Peptidase_M37; 1.
 DR SMART: SM00257; LysM; 2.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 371 AA; 40664 MW; 88800C5E5E3A82A8 CRC64;

Query Match 10.0%; Score 167; DB 16; Length 371;
 Best Local Similarity 24.9%; Pred. No. 0.00025;
 Matches 72; Conservative 47; Mismatches 122; Indels 48; Gaps 11;

QY 73 YGVKQDTSKIAQRYGLNMRREIGHINNLSSTYITYGOMLTWMSGD---LKYERSISS 129
 DB 87 YVEVEGDVVGRIAGRYDISDPAISLNKLRSTRALQVGLKIPSDGILYTVKNGDTFS 146
 QY 130 GVNTAHPSPFPAVQSSRPVQOHFAVOKFPTPPVV-VVKKP-----TPTP----- 172
 DB 147 SIAMAH-----QISLERLVLTNPSSSKESPVSRTIVSPFYNASARESCEVFPFSSAKQ 201
 QY 173 -----PVVQ--QPAVAPVTEAPFATGSSGVMQ-----FPYVQATNPVVRFRF-- 215
 DB 202 WRENTSPDAVQPLQPA-----RVLPFLPGHLSARALQELINGDLFRAPLSRYVSRYGWR 257
 QY 216 TATVAGSVTVSNGWMPFGSDGLINASAGVIGADHN-MGASIVIOHTNGFVSSTYHI 274
 DB 258 SDPTGASFPNGIDMWSRGTPTVYSAIGLVRYGVGAIVGVNLIIVHAGYQTLVYHL 317
 QY 275 KDAQVKTGDTYRTGQRIASM-KNOPSQALPEFRISNGVYVDPITVVK 322
 DB 318 QTVLVSAGTRVTSATKIGLTKRGSTGPHLFTYKNGASINPTSLK 366

RESULT 30
 Q9KJW8 PRELIMINARY; PRT; 401 AA.

AC Q9KJW8;
 DT 01-OCT-2000 (TREMblrel. 15; Created)
 DT 01-OCT-2000 (TREMblrel. 15; Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19; Last annotation update)
 DE 43 kDa antigen.
 OS *Bartonella bacilliformis*.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bartonellaceae; Bartonella.
 OX NCBI_TaxID=774;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KCS84;
 RX MEDLINE=20407302; PubMed=10948113;
 RA Padmalayan I., Kelly T., Baumstark B., Massung R.;
 RT "Molecular cloning, sequencing, expression, and characterization of an
 immunogenic 43-kDa outer lipoprotein of *Bartonella bacilliformis* that
 has homology to NLPD/LppB.";
 RL Infect. Immun. 68:4972-4979(2000).
 DR EMBL: AF157831; AA060360.1; -
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR002886; Peptidase_M37.
 DR Pfam: PF01476; LysM; 1.
 DR Pfam: PF01551; Peptidase_M37; 1.
 DR SMART: SM00257; LysM; 1.
 SQ SEQUENCE 401 AA; 43418 MW; C5ABFE0E8575FBD6 CRC64;

Query Match 9.9%; Score 165.5; DB 2; Length 401;
 Best Local Similarity 25.0%; Pred. No. 0.00036;
 Matches 83; Conservative 51; Mismatches 141; Indels 57; Gaps 14;
 QY 31 AGCASKPTYN-STSGSGS-----HRTSGSG-----GLAIGSVITDSQGVNRYOV 75

DB 86 SCDHNASVYFSSHOEAGSSFRRENKVPDGRIMGSPONIGTLRSVNNPSPIFOKGSYIV 145
 QY 76 KQDQTSKIAQRYGLNMRREIGHINNLSSTYITYGOMLTWMSGDLYKERSISSGVNTAH 135
 DB 146 OSQDITLFSIAQKIGISVESIKLVANGMDN-AIYIGQKLVIPDKQVTVTSNAONNAVGSSST 204
 QY 136 TSPFPAVQSS--SRPVOOHFAVOKP--FPVVVVVKKFPTPPVVOQPAVAPVTEAPF 190
 DB 205 TTS--STSGKMGSVFKKEEISASRSVNTAPALIN---TPDKKN---VSKOMTSSKS 255
 QY 191 ATGSS-----GVMQFPYVQATNPVVRFGTATVAGSVTVSNGWMPFGSD 235
 DB 256 ASNSKTVKDSDMITVPAQFVSKMRFPVAG-----RLSHVQKQKGTITNRGIDINVE 310
 QY 236 GDILNMSNACTVIGADHND--GASIVIOHTNGFVSSTYHIKQAVKGTGRTGRIAS 293
 DB 311 GSSVKAENGVVIVYASDGLKELGVNVRHEDNIIITIGHNSKLTYSRGRVXRGDEIA 369
 QY 294 MKNOPSQAL--FEFRISNGVYVDPITVVK 322
 DB 370 -KSGVSGDATTTPVYFEIRONSVPVNPAYLE 400

RESULT 31
 Q8Y4E2 PRELIMINARY; PRT; 436 AA.

AC Q8Y4E2;
 DT 01-MAR-2002 (TREMblrel. 20; Created)
 DT 01-MAR-2002 (TREMblrel. 20; Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21; Last annotation update)
 DE Hypothetical protein Imoz2504.
 GN IMO2504.
 OS *Listeria monocytogenes*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD5 / SEROVAR 1/2A;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Frangeul L., Buchteler C., Ruenick C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brand P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Denoux P.,
 RA Domian E., Dominguez-Bernal G., Duchaud E., Duran L., Duseigneur O.,
 RA Entian K.-D., Feihl H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hahn T., Haut U., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Ramirez-Boland J.-A., Voss H., Weiland U., Cossart P.;
 RT "Comparative genomics of *Listeria species*.";
 RL Science 294:849-852(2001).
 DR EMBL: AL591983; CAD00582.1; -
 DR ListList; LMO02504; -
 DR InterPro: IPR002886; Peptidase_M37.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF01551; Peptidase_M37; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 436 AA; 47049 MW; 2D36887C1098E167 CRC64;

Query Match 9.5%; Score 158.5; DB 16; Length 436;
 Best Local Similarity 24.8%; Pred. No. 0.0013;
 Matches 66; Conservative 33; Mismatches 100; Indels 67; Gaps 11;

QY 76 KQDQTSKIAQRYGLNMRREIGHINNLSSTYITYGOMLTWMSGDLYKERSISSGVNTAH 135
 DB 210 EKNDLVVALANKKULT-----KSEQTLAS-----QALITDEKILASINAGK 254
 QY 136 TSPFPAVQSSRPVQOHFAVOKFPTPPVVVYVKKFPTPPVVOQPAVAPVTEAPATGSS 195
 DB 255 AKQEAATKAAEKWQEAALAAASAKSAVVK-----QFSSSSNEATE-----TVSS 300

QY 196 GVAQFRYPVGA-----TNPVRRF-----GTATVAGSTVTSNGMFMFGSDGLINAS 242
 DB 301 GGGGFIIPASGILITSGSERSTNPTGKYESHKGGDIDAGGIVT-----VSA 347
 QY 243 NAGTVIOADNMDCAS-----IVIOHTNGFVSSYIHKDAQV--TGDTRTGRIAS 293
 DB 348 ASGTVPVPSGAGSGSGFGYGVKIDHNGFQTLGYHMRAGSLKVTVGQGVSGGPIGI 407
 QY 294 M--KNQSSGALPFRSRNGVYDP 317
 DB 408 MGSTGSTGQHL-HFEIRHNGVDP 432

RESULT 32

QY 08RG41 PRELIMINARY; PRT; 321 AA.
 AC 08RG41;
 DT 01-JUN-2002 (Tremblrel, 21, Created)
 DT 01-JUN-2002 (Tremblrel, 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
 DE Cell wall endopeptidase family M23/M37.
 GN FMO477.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteriia; Fusobacterium.
 OX NCBI_Taxid=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; Pubmed=11889109;
 RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman B., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 RA Fomstein M., Kyriades N., Overbeck R.,
 RA "Genome sequence and analysis of the oral bacterium Fusobacterium
 RT nucleatum strain ATCC 25586."
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL: AE010559; AAL94673.1; -
 KW Complete proteome.
 SQ SEQUENCE 321 AA; 35455 MW; 0E4158B02AA52DD9 CRC64;

Query Match 9.5%; Score 158; DB 16; Length 321;
 Best Local Similarity 24.9%; Pred. No. 0.00093;
 Matches 85; Conservative 49; Mismatches 125; Indels 82; Gaps 18;

QY 5 IAINSONQKPKIKLIGVITTCILAGCASKPTTSTSGSHRTSGSGGLAIGSVIT 64
 DB 40 IYKVKQKDDVQVAKKFGVPERIMI-----NNQFAMDNKLKVG-EVL 83
 QY 65 DSGQVPRRYGKODYSKIAQRYGLMREIGHINNINSSYITYGQWLTMGDLK--- 121
 DB 84 PPSIDGLYTLQKEMAKVAKKYGKVVVDIGDNNINPK-KLAGTTLFLKGVTLKKYK 142
 QY 122 -VERSSISGVNTAHTSPVAVQSSRPVQOHPAVOKPTPPVVVKKTPPVVQOPAP 180
 DB 143 EVEGRLL-----AAQAKKE-QKKEAQQKGG-----KKGSGAP----- 176
 QY 181 VAPPVT-----EAPFATGSGVMQFRYPV--GATNPVRRFETATVAGSTVTSNGMP 231
 DB 177 --PDDTGGDGGAGPASYSGE--FAFPYVAGITSPFGNRY--HPLKRYIHATV-- 227
 QY 232 SGRGDLI-----NASNAGTVIOADNMDCAS--IYIOHTNGFVSSYIHKDAQV 283
 DB 228 ----DLVAKYVPLRASKAGVVTAG-NMSGYKIIIIKHNGYETRAHLSTVSTNGE 281
 QY 284 TVRTGRIASMKN--QPSGALPFRSRNGVYDPPLTVLK 322
 DB 282 HVMKGDILGKTNGSGRTTGHL-HFEIRHNGVDPNPKKYLQ 321

RESULT 33

Q9RM21

ID Q9RM21 PRELIMINARY; PRT; 375 AA.
 AC Q9RM21;
 DT 01-MAY-2000 (Tremblrel, 13, Created)
 DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
 DE Lipoprotein, putative.
 GN DR0848.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_Taxid=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE=20036896; Pubmed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RA "Genome sequence of the radioreistant bacterium Deinococcus
 RT radiodurans R1."
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001939; AAF10427.1; -
 DR TIGR: DR0848;
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR002886; Peptidase_M37.
 DR Pfam: PF01476; LysM; 1.
 DR Pfam: PF01551; Peptidase_M37; 1.
 DR SMART: SM00257; LysM; 1.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 375 AA; 39107 MW; 3FF9A242F97C00AE CRC64;

Query Match 9.5%; Score 158; DB 16; Length 375;
 Best Local Similarity 21.7%; Pred. No. 0.0011;
 Matches 74; Conservative 36; Mismatches 105; Indels 126; Gaps 12;

QY 68 GVPNRVYKQGDYSKIAQRYGLMREIGHINNINSSYITYGQWLTMGDLKVRERSI 127
 DB 14 GVAQAYTVKKKGLDLYSLARGSGLTVDALMRLNGL-STPELRVGGVITL-----P 61
 QY 128 SSGVNTAHTSPVAVQSSRP-PVQOHPAVOKPT--PPV--VKKP--TPRP----- 173
 DB 62 GEGATPATPAAPATPATPAAPAPAPAVPTGLPVAANAVLKPGLTITTPKALM 121
 QY 174 -----VQOPAPVAPPVTE-----APFATGSGVMQFRYPVAGATNPV- 211
 DB 122 GGGFALRLSGEQANVTVRRPSELGEDVRQPNELRLMSGQYIVPGRVILGKTTPIY 181
 QY 212 -----RR----- 213
 DB 182 EYRLGDDVVGQIPVGDLGQRVQHLNPPSISKYLOPAREADPAVEAOAYORFTPOOMQ 241
 QY 214 --FGTATVAGSTVTSN-----GMWFGSDGLINASNAGTVIOAD-HN 253
 DB 242 QPFSALASGKATSSSGQRTYVAGKVAHYHGTIDYPAASGTAVALINQSTVIAGRYP 301
 QY 254 MDGASIVIOHTNGFVSSYIHKDAQVKTGDTVRGRIASM 294
 DB 302 VRGLVVIDHAGAVSLYFHSQKVTAKVAGQOIKRQGVGEV 342

RESULT 34

Q927Y9

ID Q927Y9 PRELIMINARY; PRT; 437 AA.
 AC Q927Y9;
 DT 01-DEC-2001 (Tremblrel, 19, Created)
 DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
 DE Hypothetical protein lin2647.
 GN LIN2647.

OS *Listeria innocua*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP_11262 / SEROVAR 6A;
 RA PubMed=11679669;
 RA Glaeser P., Fraenkel L., Buchrieser C., Rusniok C., Amend A.,
 Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 Charbit A., Checroun F., Cove E., de Darvar A., Denoux P.,
 Domian E., Dominguez-Bernal G., Duchaud E., Durand L., Dussurget O.,
 Entian K.-D., Fish H., Garcia-del Portillo F., Garrido P.,
 Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 Jones L.-M., Kaerst U., Kreft J., Kunh W., Kuntz F., Kutarpak G.,
 Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 Rasmussen G., Novella S., de Pablos B., Perez-Diaz J.-C.,
 Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.,
 "Comparative genomics of *Listeria species*."
 RA Science 294:849-852(2001).
 RL EMBL; AL596173; CAC97874.1; -
 DR Listeria; L102647; -
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR Hypothetical protein; Complete proteome.
 KM SEQUENCE 437 AA; 47301 MW; F3FD028682428C57 CRC64;
 SQ
 Query Match 9.3%; Score 155; DB 16; Length 437;
 Best Local Similarity 23.3%; Pred. No. 0.0023;
 Matches 62; Conservative 35; Mismatches 10; Indels 66; Gaps 10;
 QY 76 KQGTYSKIAORVGLNMRREIHNNSYTYTGOMLTMSGLKXREBSISGVNTH 135
 DB 210 EKNLVVALAKVKDIT-----KSEQTLITNE-----QGLSDQKCLAKINAEK 254
 QY 136 TPSPVAVOSSRPVQOHPAYQKPPPPVVKPPPTPPVVOQPAVAPVTEAPFATGSS 195
 DB 255 AKQENAIKAAEKKRQENAAKNAT-----VAAQOPSSVTSSAGTATDTVSS 301
 QY 196 GVMQFRYPVGA-----TNPVVRP-----GTATVAGSTVTSNGMWSGRDGLINAS 242
 DB 302 GCGGFIRKASQMLTSGFSERTNTPVKGKESHKQDIAAGGIT-----VSAA 348
 QY 243 NAGTVIQADHNMMDGAS-----IVIQHTNGFVSSYIHIKDAQVK--TQDVTYRTGORIAS 293
 DB 349 ASGRVVFSGFATSGSGFGYGVVVIDHNGFQTLYAHMRAGSLKVTGQGVSGOPIGI 408
 QY 294 M--KNQPSGALFEPRIISNGVYVDP 317
 DB 409 MGSTGQSTGQHL-HPEIKHNGIPVDP 433
 RESULT 35
 OS2606 PRELIMINARY; PRT; 176 AA.
 ID OS2606
 AC 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Novel 11poprotein N1PD (Fragment).
 GN N1PD.
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AW1;
 RX MEDLINE=98294041; PubMed=9632252;
 RA Flavier A.B., Schell M.A., Denny T.P.;
 "An Rpos (sigmas) homologue regulates acylhomoserine lactone-dependent

RT autoinduction in *Ralstonia solanacearum*.
 RL Mol. Microbiol. 28:475-486(1998).
 DR EMBL; AF042351; AAC38576.1; -
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01551; Peptidase_M37; 1.
 KM Lipoprotein.
 FT NON TER
 SQ SEQUENCE 176 AA; 18450 MW; EBB5214BAEA6300 CRC64;
 Query Match 9.2%; Score 153.5; DB 2; Length 176;
 Best Local Similarity 28.8%; Pred. No. 0.0009;
 Matches 51; Conservative 26; Mismatches 79; Indels 21; Gaps 6;
 QY 153 PAVQKPPPPVYVKKPPPTPP--VVOQPAVAPVTEAPFATGSS-----GVMQFRYPVG 205
 DB 10 FNADANTNGVVV-----TPVRPGNTTSDPITRAVTSPPVVASGASAGATGALALAMP-- 63
 QY 206 ATNPVVRPPTATVAGSTVTSNGMWSGRDGLINASNAGTVIQADHND--GASIVYQHT 264
 DB 64 AHQGVIGRF-----DDKANKGIDIGKKRGDPVSAADGKVIHVGFLRGYGNLVITKHN 116
 QY 265 NGFVSSYIHKDAQVKTGDTVTRTGORIASMNQPSGAALFEPRIISNGVYVDPITVL 321
 DB 117 DTELTAYGNNDKVLVTEOSTVTKGQKLAEMGSTDADRVKLFVVRNGKVPDWRFL 173
 RESULT 36
 OS2606 PRELIMINARY; PRT; 228 AA.
 ID OS2606
 AC 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative Peptidase.
 GN SC03168 OR SCE94.19C.
 OS *Streptomyces coelicolor*.
 OC Actinobacteria; Actinobacteriales; Streptomyces.
 OC Actinobacteria; Streptomyces.
 NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Oliver K., Harris D.;
 Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D.; Parkhill J.; Barrell B.G.; Rajandream M.A.;
 Bentley S.D.; Parkhill J.; Barrell B.G.; Rajandream M.A.;
 Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapate D., Richner A., Cullum J.,
 Kinasl H., Hopwood D.A.;
 "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb *Streptomyces coelicolor* A3(2) chromosome."
 RT Mol. Microbiol. 21:77-96(1996).
 RL [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 Rabinowitsch E., Rajandream M.A., Rutherford R., Rutter S.,
 Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
 Hopwood D.A.;
 "Complete genome sequence of the model actinomycete *Streptomyces*
 coelicolor A3(2)."
 RT Nature 417:141-147(2002).

DR EMBL; AL049628; CAB40868.1; -
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
SQ SEQUENCE 228 AA; 22817 MW; 2263DF3167B0D279 CRC64;

Query Match 9.2%; Score 153; DB 16; Length 228;
Best Local Similarity 37.7%; Pred. No. 0.0014;
Matches 46; Conservative 16; Mismatches 42; Indels 18; Gaps 6;

QY 219 VAGSTVSNQMW---FSGRD-----GDLINASNAGTYIADHN-----MGASIVIOHT 264
DB 101 LSASFQNGGMMHAKSGQDPVAVIGTIVVAHGGTIVKAGGNGAGDGPAYGMAIVIKHG 160
QY 265 NGFVSSYIHKDAQVKTGDTYRTGQRIASMKR--QPSGAL--FEFRISRN--GVYVDPPLTV 320
DB 161 NGTSSQYAHLSRIWVKIGQIVKTQSTAKSGNTGNSGPHLFEIRTPNVGSAVDPAVF 220
QY 321 LK 322
DB 221 LR 222

RESULT 37
ID Q9X6S4 PRELIMINARY; PRT; 203 AA.
AC Q9X6S4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Lipoprotein NlpD (Fragment).
GN NLPD.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB22;
RA McCarter L.L.;
RT "RNA polymerase sigma factor Rpos of Vibrio parahaemolyticus";
RL EMBL; AF144608; AAD33932.1; -
DR EMBL; AF144608; AAD33932.1; -
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
FT NON TER 1
SQ SEQUENCE 203 AA; 21915 MW; 33EB174EADFA2630 CRC64;

Query Match 9.1%; Score 152.5; DB 2; Length 203;
Best Local Similarity 26.4%; Pred. No. 0.0013;
Matches 56; Conservative 34; Mismatches 75; Indels 47; Gaps 9;

QY 137 PSPVAVQSSRPVQGHVAVOKP---TPPVVVVK---KPPPTPVVQGPAP----- 180
DB 3 PVPVPPVTTTP---PVPVAKFTTTTKPVTSKSSQKPTTTPKVAQKEBPKEVEQTKA 57
QY 181 -----VAPVTEAPFATGSSGWMQFRYPVAGATPVVRRGATVAVASTVTSN 227
DB 58 KEYVSGKDNQHTKRPPTT---TVQNDKYSKMLMP---TKGRVAKNS---AEQ--GNK 106
QY 228 GMMFSGDGLINASNAGTYIADHNMDGAS--IVIOHTNGFVSSYIHKDAQVKTGDTV 285
DB 107 GIDVAGRGQPIVSTAAGTIVVYSGNALRGVNLIIIVKNDVYLSAVAHNDGLVTEQSQSV 166
QY 286 RTGORIASMKQPSGALFEFRISRNQVYVDP 317
DB 167 KSGQKLTMGSSGAKSVKLFHFRYQGSVNP 198

RESULT 38
ID Q8YRUO PRELIMINARY; PRT; 760 AA.
AC Q8YRUO;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein A1R3353.
GN A1R3353.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=1036590;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsuno M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003592; BAB75052.1; -
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 760 AA; 80569 MW; FB62A0DDFCA573B CRC64;

Query Match 9.1%; Score 152; DB 16; Length 760;
Best Local Similarity 20.1%; Pred. No. 0.0079;
Matches 95; Conservative 41; Mismatches 112; Indels 224; Gaps 18;

QY 61 QVITD--SQGVNRYQVKGQDTSKIAQRYGLNREIGHNNLNSSYTYTGOWTLWSGD 119
DB 294 QPLTETSASNTAYEVVPGDTLLAASRVNTSVAVELVKVNNLSNPDLKXISQQLIIS-- 351
QY 120 LKVRNSISGCVNTATPSS-----PVAQSSRP--PVQGHVAVOK-- 157
DB 352 -----AIDSSSTIAQPAIISNRVQPGTPTVPVNPVDTARVNPFLPVAQPPSIANN 405
QY 158 -----PTP-----PV-----PVAQSSRP--PVQGHVAVOK-- 168
DB 406 SVPTVFTPGADSETPTDTIVPLESASAPAEFGVGVNVPDPAFLIQRPOQPKCAAR 465
QY 169 -----TPPPPVVQ----- 177
DB 466 AKGDRLRSLQAEIQRQAQKRDQGTGTPVAVANQNNNTAIPVTSNNFAVTRPISR 525
QY 178 -----PAVAPPTVAPATGSSGVMQFR--YPVGAT-----NPPVRRFGTATV 219
DB 526 QOEIAVPIAVPTPLAPDSNHPVKPQFRATLPVNEALNPEFLPNNAFQNPSSVRVATPPV 585
QY 220 -----AGSTV-----TSNG 228
DB 586 RLNASESLGMRKGTVPSPKLPPLAANDQYLPQALIDESVPPSDSTTAFFWPANGVLTSGY 645
QY 229 MWFSGR-----DGLINASNAGTYIADHNMDGAS--IVIOHTNGFVSSYIHKD 276
DB 646 GMMFSGDGLINASNAGTYIADHNMDGAS--IVIOHTNGFVSSYIHKD 276
QY 277 AQVKTGDTYRTGQRIASMKQPSGALF-----EFRISRNQV--YVDPPLTVL 321
DB 706 LTVQAGQVAVQGOQIAM-----GSTGFTSGPTHFRIHTKGALNPIMLV 752

RESULT 39
ID Q9FTX9 PRELIMINARY; PRT; 176 AA.
AC Q9FTX9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Lipoprotein (Fragment).
GN NLPD.
OS Escherichia coli.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:08:28 ; Search time 72 Seconds
(without alignments)
595.926 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 322
Sequence: 1 MVTVAINSQNGKRIKRLGL.....LFEFRISRGNGVYDPLTLK 322

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_101002.*

1: /SID82/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID82/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	322	100.0	322	22	AA820106
2	221	68.6	322	22	AA820107
3	9	2.8	968	22	AB861691
4	8	2.5	223	22	AA930076
5	8	2.5	557	20	AAW94997
6	7	2.2	13	18	AAW31448
7	7	2.2	13	23	AAU11458
8	7	2.2	35	22	AAE10936
9	7	2.2	54	23	ABP04087
10	7	2.2	66	22	AAW83197

11	7	2.2	67	20	AAV41355	Human secreted pro
12	7	2.2	72	22	AAU61587	Proprionbacterium
13	7	2.2	80	22	AAU50707	Proprionbacterium
14	7	2.2	90	21	AAV96225	Arabisdopsis ltrc1 c
15	7	2.2	101	23	AB855442	Lactococcus lactis
16	7	2.2	108	21	AA658104	Arabisdopsis thalia
17	7	2.2	112	22	AB866134	Drosophila melanog
18	7	2.2	114	23	ABP00920	Human ORFX protein
19	7	2.2	126	22	ABG04226	Novel human diagno
20	7	2.2	136	22	AA005078	Human polypeptide
21	7	2.2	138	21	AA658103	Arabisdopsis thalia
22	7	2.2	147	22	AA005875	Human polypeptide
23	7	2.2	157	22	AAE02227	Staphylococcus aur
24	7	2.2	159	22	AAU63705	Proprionbacterium
25	7	2.2	159	22	AAU43362	Proprionbacterium
26	7	2.2	179	23	ABP25478	Streptococcus poly
27	7	2.2	185	21	AAV75633	Neisseria gonorrhe
28	7	2.2	195	22	AB861524	Drosophila melanog
29	7	2.2	208	19	AAW71722	Arabisdopsis leafy-
30	7	2.2	208	21	AAV54563	Amino acid sequenc
31	7	2.2	208	22	AAE09838	leafy cotyledon 1
32	7	2.2	213	22	AB869336	Drosophila melanog
33	7	2.2	251	22	AAU19650	Human novel extrac
34	7	2.2	251	23	ABP47870	Human polypeptide
35	7	2.2	298	23	ABP30505	Streptococcus poly
36	7	2.2	300	21	AB808497	Envelope 2 protein
37	7	2.2	304	17	AA898460	Endonuclease Ilati
38	7	2.2	304	17	AAW02163	Lactococcus lactis
39	7	2.2	321	23	ABP28875	Streptococcus poly
40	7	2.2	337	21	AA638356	Arabisdopsis thalia
41	7	2.2	343	21	AA638355	Arabisdopsis thalia
42	7	2.2	344	19	AAW69752	Acetobacter xyliu
43	7	2.2	355	19	AAV10975	H. pylori ORF hp5p
44	7	2.2	356	22	AA883221	Yeast FARP signatu
45	7	2.2	357	21	AA609523	Arabisdopsis thalia
46	7	2.2	361	22	AB868291	Drosophila melanog
47	7	2.2	363	21	AA609522	Arabisdopsis thalia
48	7	2.2	387	21	AAV93670	Amino acid sequenc
49	7	2.2	387	21	AAV93673	Amino acid sequenc
50	7	2.2	387	21	AAV93674	Amino acid sequenc
51	7	2.2	387	21	AAV93675	Amino acid sequenc
52	7	2.2	387	22	AAV46854	S. cerevisiae RPL3
53	7	2.2	392	22	AB869425	Drosophila melanog
54	7	2.2	395	22	AB869583	Drosophila melanog
55	7	2.2	395	22	ABG10252	Novel human diagno
56	7	2.2	396	22	AB860029	Drosophila melanog
57	7	2.2	400	23	ABP29950	Streptococcus poly
58	7	2.2	401	21	AA638354	Arabisdopsis thalia
59	7	2.2	421	21	AA609521	Arabisdopsis thalia
60	7	2.2	427	22	AAU51308	Proprionbacterium
61	7	2.2	452	21	AA620779	Arabisdopsis thalia
62	7	2.2	452	21	AA638868	Arabisdopsis thalia
63	7	2.2	460	21	AA620778	Arabisdopsis thalia
64	7	2.2	460	21	AA638867	Arabisdopsis thalia
65	7	2.2	468	21	AA620777	Arabisdopsis thalia
66	7	2.2	468	21	AA638866	Arabisdopsis thalia
67	7	2.2	499	23	AB847977	Listeria monocytog
68	7	2.2	501	17	AAW03760	Mullerian inhibiti
69	7	2.2	505	14	AAK41921	MISR2A/MISR2B. Ra
70	7	2.2	505	15	AA855369	Human Activin rece
71	7	2.2	505	15	AA855373	Mouse Activin rece
72	7	2.2	505	16	AA870240	Serine/threonine k
73	7	2.2	505	20	AAV33302	Human hMLK-4 clone
74	7	2.2	505	20	AAV33306	Human hMLK-4 clone
75	7	2.2	532	22	AB864136	Drosophila melanog
76	7	2.2	532	22	AB866951	Drosophila melanog
77	7	2.2	543	18	AAW07702	Mouse FRS2 repres
78	7	2.2	571	22	AB862679	Drosophila melanog
79	7	2.2	623	20	AAV14967	Amino acid sequenc
80	7	2.2	623	22	AA883267	Yeast FARP SEQ ID
81	7	2.2	623	22	AA883275	Yeast FARP SEQ ID
82	7	2.2	674	22	ABG12515	Novel human diagno
83	7	2.2	687	21	AAV69165	A human N-acetylgl

84	7	2.2	727	22	ABG25530	Novel human diagno	157	6	1.9	14	23	AAU76046	Synthetic linker m
85	7	2.2	732	22	ABG10453	Novel human diagno	158	6	1.9	15	16	AAU79629	Endocarditis speci
86	7	2.2	753	23	ABP25819	Streptococcus poly	159	6	1.9	15	16	AAU79626	Endocarditis speci
87	7	2.2	761	23	ABP27624	Streptococcus poly	160	6	1.9	16	23	ABG62289	Endocarditis speci
88	7	2.2	765	21	AAV69167	A mature human N-a	161	6	1.9	16	23	ABG62289	Endocarditis speci
89	7	2.2	765	21	AAV69168	A mature human N-a	162	6	1.9	16	23	ABG62289	Endocarditis speci
90	7	2.2	765	21	AAV69168	A mature human N-a	163	6	1.9	16	23	ABG62289	Endocarditis speci
91	7	2.2	806	22	AAU92136	Protonibacterium	164	6	1.9	16	23	ABG62289	Endocarditis speci
92	7	2.2	814	23	ABG92136	C glutamylum prote	165	6	1.9	16	23	ABG62289	Endocarditis speci
93	7	2.2	836	22	ABG67517	Herbicidally activ	166	6	1.9	16	23	ABG62289	Endocarditis speci
94	7	2.2	842	22	ABG66631	Drosophila melanog	167	6	1.9	16	23	ABG62289	Endocarditis speci
95	7	2.2	843	20	AAV06561	Chicken capsacin	168	6	1.9	16	23	ABG62289	Endocarditis speci
96	7	2.2	843	20	AAV06561	Chicken capsacin	169	6	1.9	16	23	ABG62289	Endocarditis speci
97	7	2.2	852	22	ABG64019	Drosophila melanog	170	6	1.9	16	23	ABG62289	Endocarditis speci
98	7	2.2	864	22	ABG71319	Drosophila melanog	171	6	1.9	16	23	ABG62289	Endocarditis speci
99	7	2.2	922	21	AAV36515	Candida albicans C	172	6	1.9	16	23	ABG62289	Endocarditis speci
100	7	2.2	955	22	ABG71635	Drosophila melanog	173	6	1.9	16	23	ABG62289	Endocarditis speci
101	7	2.2	960	22	AAU75476	S. aureus antigeni	174	6	1.9	16	23	ABG62289	Endocarditis speci
102	7	2.2	1047	22	ABG59191	Drosophila melanog	175	6	1.9	16	23	ABG62289	Endocarditis speci
103	7	2.2	1062	22	ABG71830	Drosophila melanog	176	6	1.9	16	23	ABG62289	Endocarditis speci
104	7	2.2	1336	23	AAU92973	Arabidopsis transg	177	6	1.9	16	23	ABG62289	Endocarditis speci
105	7	2.2	1485	21	AAV69166	A mature human N-a	178	6	1.9	16	23	ABG62289	Endocarditis speci
106	7	2.2	2071	22	AAE10949	Mouse maeai protei	179	6	1.9	16	23	ABG62289	Endocarditis speci
107	7	2.2	2344	22	AAE10924	Staphylococcus aur	180	6	1.9	16	23	ABG62289	Endocarditis speci
108	7	2.2	2780	22	AAE10924	Staphylococcus aur	181	6	1.9	16	23	ABG62289	Endocarditis speci
109	7	2.2	2870	21	AAV95553	Caenorhabditis ele	182	6	1.9	16	23	ABG62289	Endocarditis speci
110	7	2.2	3010	21	AAV95553	Caenorhabditis ele	183	6	1.9	16	23	ABG62289	Endocarditis speci
111	7	2.2	3070	21	AAV95553	Caenorhabditis ele	184	6	1.9	16	23	ABG62289	Endocarditis speci
112	7	2.2	3178	21	AAV95553	Caenorhabditis ele	185	6	1.9	16	23	ABG62289	Endocarditis speci
113	7	2.2	3178	21	AAV95553	Caenorhabditis ele	186	6	1.9	16	23	ABG62289	Endocarditis speci
114	7	2.2	3178	21	AAV95553	Caenorhabditis ele	187	6	1.9	16	23	ABG62289	Endocarditis speci
115	7	2.2	3178	21	AAV95553	Caenorhabditis ele	188	6	1.9	16	23	ABG62289	Endocarditis speci
116	7	2.2	3178	21	AAV95553	Caenorhabditis ele	189	6	1.9	16	23	ABG62289	Endocarditis speci
117	7	2.2	3178	21	AAV95553	Caenorhabditis ele	190	6	1.9	16	23	ABG62289	Endocarditis speci
118	7	2.2	3178	21	AAV95553	Caenorhabditis ele	191	6	1.9	16	23	ABG62289	Endocarditis speci
119	7	2.2	3178	21	AAV95553	Caenorhabditis ele	192	6	1.9	16	23	ABG62289	Endocarditis speci
120	7	2.2	3178	21	AAV95553	Caenorhabditis ele	193	6	1.9	16	23	ABG62289	Endocarditis speci
121	7	2.2	3178	21	AAV95553	Caenorhabditis ele	194	6	1.9	16	23	ABG62289	Endocarditis speci
122	7	2.2	3178	21	AAV95553	Caenorhabditis ele	195	6	1.9	16	23	ABG62289	Endocarditis speci
123	7	2.2	3178	21	AAV95553	Caenorhabditis ele	196	6	1.9	16	23	ABG62289	Endocarditis speci
124	7	2.2	3178	21	AAV95553	Caenorhabditis ele	197	6	1.9	16	23	ABG62289	Endocarditis speci
125	7	2.2	3178	21	AAV95553	Caenorhabditis ele	198	6	1.9	16	23	ABG62289	Endocarditis speci
126	7	2.2	3178	21	AAV95553	Caenorhabditis ele	199	6	1.9	16	23	ABG62289	Endocarditis speci
127	7	2.2	3178	21	AAV95553	Caenorhabditis ele	200	6	1.9	16	23	ABG62289	Endocarditis speci
128	7	2.2	3178	21	AAV95553	Caenorhabditis ele	201	6	1.9	16	23	ABG62289	Endocarditis speci
129	7	2.2	3178	21	AAV95553	Caenorhabditis ele	202	6	1.9	16	23	ABG62289	Endocarditis speci
130	7	2.2	3178	21	AAV95553	Caenorhabditis ele	203	6	1.9	16	23	ABG62289	Endocarditis speci
131	7	2.2	3178	21	AAV95553	Caenorhabditis ele	204	6	1.9	16	23	ABG62289	Endocarditis speci
132	7	2.2	3178	21	AAV95553	Caenorhabditis ele	205	6	1.9	16	23	ABG62289	Endocarditis speci
133	7	2.2	3178	21	AAV95553	Caenorhabditis ele	206	6	1.9	16	23	ABG62289	Endocarditis speci
134	7	2.2	3178	21	AAV95553	Caenorhabditis ele	207	6	1.9	16	23	ABG62289	Endocarditis speci
135	7	2.2	3178	21	AAV95553	Caenorhabditis ele	208	6	1.9	16	23	ABG62289	Endocarditis speci
136	7	2.2	3178	21	AAV95553	Caenorhabditis ele	209	6	1.9	16	23	ABG62289	Endocarditis speci
137	7	2.2	3178	21	AAV95553	Caenorhabditis ele	210	6	1.9	16	23	ABG62289	Endocarditis speci
138	7	2.2	3178	21	AAV95553	Caenorhabditis ele	211	6	1.9	16	23	ABG62289	Endocarditis speci
139	7	2.2	3178	21	AAV95553	Caenorhabditis ele	212	6	1.9	16	23	ABG62289	Endocarditis speci
140	7	2.2	3178	21	AAV95553	Caenorhabditis ele	213	6	1.9	16	23	ABG62289	Endocarditis speci
141	7	2.2	3178	21	AAV95553	Caenorhabditis ele	214	6	1.9	16	23	ABG62289	Endocarditis speci
142	7	2.2	3178	21	AAV95553	Caenorhabditis ele	215	6	1.9	16	23	ABG62289	Endocarditis speci
143	7	2.2	3178	21	AAV95553	Caenorhabditis ele	216	6	1.9	16	23	ABG62289	Endocarditis speci
144	7	2.2	3178	21	AAV95553	Caenorhabditis ele	217	6	1.9	16	23	ABG62289	Endocarditis speci
145	7	2.2	3178	21	AAV95553	Caenorhabditis ele	218	6	1.9	16	23	ABG62289	Endocarditis speci
146	7	2.2	3178	21	AAV95553	Caenorhabditis ele	219	6	1.9	16	23	ABG62289	Endocarditis speci
147	7	2.2	3178	21	AAV95553	Caenorhabditis ele	220	6	1.9	16	23	ABG62289	Endocarditis speci
148	7	2.2	3178	21	AAV95553	Caenorhabditis ele	221	6	1.9	16	23	ABG62289	Endocarditis speci
149	7	2.2	3178	21	AAV95553	Caenorhabditis ele	222	6	1.9	16	23	ABG62289	Endocarditis speci
150	7	2.2	3178	21	AAV95553	Caenorhabditis ele	223	6	1.9	16	23	ABG62289	Endocarditis speci
151	7	2.2	3178	21	AAV95553	Caenorhabditis ele	224	6	1.9	16	23	ABG62289	Endocarditis speci
152	7	2.2	3178	21	AAV95553	Caenorhabditis ele	225	6	1.9	16	23	ABG62289	Endocarditis speci
153	7	2.2	3178	21	AAV95553	Caenorhabditis ele	226	6	1.9	16	23	ABG62289	Endocarditis speci
154	7	2.2	3178	21	AAV95553	Caenorhabditis ele	227	6	1.9	16	23	ABG62289	Endocarditis speci
155	7	2.2	3178	21	AAV95553	Caenorhabditis ele	228	6	1.9	16	23	ABG62289	Endocarditis speci
156	7	2.2	3178	21	AAV95553	Caenorhabditis ele	229	6	1.9	16	23	ABG62289	Endocarditis speci

230	6	1.9	23	23	ABR78497	Gum arabic glycop	303	6	1.9	67	21	AA602271	Human secreted pro
231	6	1.9	24	20	AAV01271	Gum arabic glycop	304	6	1.9	68	22	AAU48057	Proionbacterium
232	6	1.9	24	20	ABR78498	Gum arabic glycop	305	6	1.9	69	22	AAU50589	Proionbacterium
233	6	1.9	32	21	AAV95780	Glycosylphosphidyl	306	6	1.9	69	22	AAW1887	Human bone marrow
234	6	1.9	32	22	ABR23454	Protein #5453 enco	307	6	1.9	69	23	ABG41700	Human peptide enco
235	6	1.9	32	22	AAW58895	Human brain expres	308	6	1.9	70	22	AAU57491	Proionbacterium
236	6	1.9	32	22	AAW1421	Human bone marrow	309	6	1.9	70	22	AAU60090	Proionbacterium
237	6	1.9	32	22	AAW1067	Human peptide enco	310	6	1.9	70	22	ABB10629	Human pancreatic c
238	6	1.9	32	23	ABG41225	Human peptide enco	311	6	1.9	70	22	AAW87429	Human immune/haema
239	6	1.9	33	22	ABR38493	Protein #5399 enco	312	6	1.9	70	22	AAW92385	Human digestive sy
240	6	1.9	33	22	ABR23645	Protein #5644 enco	313	6	1.9	71	22	AAU63040	Proionbacterium
241	6	1.9	33	22	AAW59109	Human brain expres	314	6	1.9	72	22	AAU50948	Proionbacterium
242	6	1.9	33	22	AAW1644	Human bone marrow	315	6	1.9	72	22	AAU66354	Proionbacterium
243	6	1.9	33	22	AAW19248	Peptide #5682 enco	316	6	1.9	73	22	AAU41987	Proionbacterium
244	6	1.9	33	22	AAW1938	Peptide #5975 enco	317	6	1.9	74	22	AAU63140	Proionbacterium
245	6	1.9	33	23	ABG41458	Human peptide enco	318	6	1.9	74	22	AAW92770	Human excretory re
246	6	1.9	34	21	ABR15436	TCR beta V-N-J reg	319	6	1.9	74	22	AAW42585	Human kidney relat
247	6	1.9	34	22	AAO13463	Human polypeptide	320	6	1.9	74	23	ABR78446	Gum arabic glycop
248	6	1.9	34	22	AAW77262	Human colon cancer	321	6	1.9	76	20	AAW98100	Yeast cdc24 matng
249	6	1.9	38	21	AAV93764	Amino acid sequenc	322	6	1.9	76	22	AAU54993	Proionbacterium
250	6	1.9	39	23	ABR78494	Gum arabic glycop	323	6	1.9	76	22	ABR43310	Peptide #10816 enc
251	6	1.9	39	23	ABR78445	Gum arabic glycop	324	6	1.9	76	22	AAW64218	Human brain expres
252	6	1.9	41	23	ABR81352	Human prostate spe	325	6	1.9	76	22	AAW77043	Human bone marrow
253	6	1.9	42	21	ABR53853	Human colon cancer	326	6	1.9	76	22	AAW37157	Peptide #1194 enc
254	6	1.9	43	20	AAV14015	Yaah repeat sequen	327	6	1.9	76	23	ABP07349	Human ORFX protein
255	6	1.9	44	20	AAV13971	LysB repeat sequen	328	6	1.9	77	21	AAW98883	Maize embryo spect
256	6	1.9	44	20	AAV13973	XyIA repeat sequen	329	6	1.9	77	21	AAW32894	Arabidopsis thalia
257	6	1.9	44	20	AAV13969	Lys repeat sequenc	330	6	1.9	77	22	AAW85089	Human immune/haema
258	6	1.9	44	22	ABR50846	Human secreted pro	331	6	1.9	79	17	AAW97976	Fusion protein car
259	6	1.9	45	23	ABP30650	Streptococcus poly	332	6	1.9	80	23	ABP07845	Human ORFX protein
260	6	1.9	48	22	ABR42215	Peptide #9721 enco	333	6	1.9	81	22	ABR57836	Drosophila melanog
261	6	1.9	48	22	AAW63100	Human brain expres	334	6	1.9	82	22	AAU42827	Proionbacterium
262	6	1.9	48	22	AAW5911	Human bone marrow	335	6	1.9	82	23	ABP26938	Streptococcus poly
263	6	1.9	48	22	AAW36022	Peptide #10059 enc	336	6	1.9	83	21	AAW33785	Arabidopsis thalia
264	6	1.9	49	19	AAW59911	Amino acid sequenc	337	6	1.9	83	22	AAW86534	Novel human connec
265	6	1.9	50	22	AAU57285	Proionbacterium	338	6	1.9	83	22	AAU47318	Proionbacterium
266	6	1.9	50	22	AAU53205	Proionbacterium	339	6	1.9	83	23	ABP1716	Human ORF689 prote
267	6	1.9	51	23	ABP35431	Human ORF4404 prot	340	6	1.9	84	22	AAU20521	Human secreted pro
268	6	1.9	51	22	AAU64877	Proionbacterium	341	6	1.9	85	22	ABG06522	Novel human diagno
269	6	1.9	52	17	AAW02093	S. mutans antigen	342	6	1.9	85	23	ABP00054	Human ORFX protein
270	6	1.9	53	22	AAU57802	Proionbacterium	343	6	1.9	86	21	AAW6563	C. elegans insulin
271	6	1.9	54	20	AAU02724	Human secreted pro	344	6	1.9	86	22	AAU61512	Proionbacterium
272	6	1.9	54	23	ABP34838	Human ORF3811 prot	345	6	1.9	87	22	ABR68252	Drosophila melanog
273	6	1.9	55	22	ABR41095	Peptide #8601 enco	346	6	1.9	89	23	ABR49133	Listeria monocytog
274	6	1.9	55	22	ABR25148	Protein #7147 enco	347	6	1.9	92	21	AAW01138	Human secreted pro
275	6	1.9	55	22	AAW61951	Human brain expres	348	6	1.9	92	22	AAW51102	Proionbacterium
276	6	1.9	55	22	AAW74755	Human bone marrow	349	6	1.9	92	22	AAW97226	Melanoma polypep
277	6	1.9	55	22	AAO08095	Human polypeptide	350	6	1.9	93	22	AAU27941	Human contig polyp
278	6	1.9	55	22	AAW20401	Peptide #6835 enco	351	6	1.9	93	22	AAO09387	Human polypeptide
279	6	1.9	55	22	AAW44871	Peptide #8908 enco	352	6	1.9	93	23	AAU96807	Cancer gene therap
280	6	1.9	55	23	ABG44552	Human peptide enco	353	6	1.9	94	22	AAO08047	Human polypeptide
281	6	1.9	56	22	ABP05095	Human ORFX protein	354	6	1.9	95	21	AAW57203	Arabidopsis thalia
282	6	1.9	56	22	AAU55761	Proionbacterium	355	6	1.9	95	21	AAW59156	Arabidopsis thalia
283	6	1.9	56	22	AAW83969	Human immune/haema	356	6	1.9	96	22	AAU63242	Proionbacterium
284	6	1.9	58	22	AAU67225	Proionbacterium	357	6	1.9	97	18	AAW44126	Streptococcus pneu
285	6	1.9	59	22	AAU54245	Proionbacterium	358	6	1.9	97	22	AAU47608	Proionbacterium
286	6	1.9	59	22	ABG33751	Novel human diagno	359	6	1.9	97	22	AAW66331	Proionbacterium
287	6	1.9	60	22	AAU39971	Proionbacterium	360	6	1.9	97	23	ABP47813	WNV protein relat
288	6	1.9	60	22	AAO13850	Human polypeptide	361	6	1.9	98	21	AAW15079	Arabidopsis thalia
289	6	1.9	61	22	AAW88541	Human immune/haema	362	6	1.9	98	22	AAW83809	Human immune/haema
290	6	1.9	61	23	ABP08454	Human ORFX protein	363	6	1.9	100	22	AAU59584	Proionbacterium
291	6	1.9	62	22	AAU52735	Proionbacterium	364	6	1.9	100	23	ABG59960	Human DTHP polype
292	6	1.9	62	22	AAO11568	Human polypeptide	365	6	1.9	101	21	AAW55982	Arabidopsis thalia
293	6	1.9	62	22	AAW5773	Human colon cancer	366	6	1.9	101	22	AAW57377	Proionbacterium
294	6	1.9	63	22	AAU39814	Proionbacterium	367	6	1.9	101	22	AAW92336	C glutaminc prote
295	6	1.9	63	22	AAU51768	Proionbacterium	368	6	1.9	102	19	AAW58889	S. pneumoniae deri
296	6	1.9	63	22	AAW92833	C glutaminc prote	369	6	1.9	102	22	AAW06407	Human foetal prote
297	6	1.9	64	21	AAW25878	Zea mays protein f	370	6	1.9	103	22	AAW86912	Human immune/haema
298	6	1.9	64	22	AAU51537	Proionbacterium	371	6	1.9	103	22	AAW72557	Escherichia coli s
299	6	1.9	64	22	AAU64736	Proionbacterium	372	6	1.9	104	21	AAW08445	A human prostate s
300	6	1.9	65	20	AAW88927	Sequence ID #527 f	373	6	1.9	104	21	AAW34476	Arabidopsis thalia
301	6	1.9	65	22	ABW50956	Human secreted pro	374	6	1.9	104	21	AAW45464	Arabidopsis thalia
302	6	1.9	66	22	AAW82784	Human immune/haema	375	6	1.9	104	22	AAU42872	Proionbacterium

376	6	1.9	105	22	AAU50456	Proionbacterium
377	6	1.9	105	22	ABR29071	Peptide #1722 enco
378	6	1.9	105	22	ABR34231	Peptide #1739 enco
379	6	1.9	105	22	ABR50792	Human secreted pro
380	6	1.9	105	22	AAW5027	Human brain exores
381	6	1.9	105	22	AAW57412	Human bone marrow
382	6	1.9	105	22	AAW5335	Human immune/haema
383	6	1.9	105	22	AAW56807	Peptide #1577 enco
384	6	1.9	105	22	AAW12543	Peptide #1741 enco
385	6	1.9	105	22	AAW27704	Peptide #1668 enco
386	6	1.9	105	22	AAW29986	Human cancer assoc
387	6	1.9	106	21	AAW43541	Alpha-amylase sequ
388	6	1.9	106	22	ABP60827	Manduca sexta chlo
389	6	1.9	106	23	ABP60827	Human anti-RSV mon
390	6	1.9	107	18	AAW1634	Escherichia coli p
391	6	1.9	107	22	ABW52872	Human foetal prote
392	6	1.9	108	22	ABW58598	Drosophila melanog
393	6	1.9	108	22	ABW58598	Drosophila melanog
394	6	1.9	108	22	ABW58598	Human nCPER-Seg107
395	6	1.9	109	23	ABW58598	Arabidopsis thalia
396	6	1.9	110	21	AAW16254	Arabidopsis thalia
397	6	1.9	110	21	AAW16254	Proionbacterium
398	6	1.9	110	22	AAW60794	Proionbacterium
399	6	1.9	111	23	ABP34224	Human ORF3197 prot
400	6	1.9	111	23	ABP34224	Plinus radiata tran
401	6	1.9	115	21	ABP34280	Human ovariian anti
402	6	1.9	115	23	ABP42590	MSP1-derived poly
403	6	1.9	116	16	AAW80187	Human nervous syst
404	6	1.9	116	16	AAW80189	Human digestive sy
405	6	1.9	116	22	ABW17128	Human liver antiq
406	6	1.9	116	22	ABW19271	Human liver antiq
407	6	1.9	116	22	AAU20067	Human ORFX protein
408	6	1.9	116	22	ABP40928	Human ORFX protein
409	6	1.9	116	23	ABP09847	Human ORFX protein
410	6	1.9	116	23	ABP1431	Novel human digno
411	6	1.9	117	22	ABG13570	Human polypeptide
412	6	1.9	117	22	AAO17337	Human secreted pro
413	6	1.9	119	21	AAW08893	Proionbacterium
414	6	1.9	119	22	AAW48739	Human ovariian anti
415	6	1.9	120	22	AAU51633	Arabidopsis thalia
416	6	1.9	120	23	ABP41143	Proionbacterium
417	6	1.9	121	21	AAW34475	Human protein sequ
418	6	1.9	121	22	AAW55052	Human prostate can
419	6	1.9	122	22	AAW55052	H. pylori ORF 299e
420	6	1.9	123	21	AAW57118	Human pancreatic c
421	6	1.9	124	18	AAW55414	Plinus radiata tran
422	6	1.9	124	21	AAW54240	Human ORF62 prote
423	6	1.9	124	21	ABP33267	Novel human digno
424	6	1.9	124	23	ABP31689	Human polypeptide
425	6	1.9	125	22	AAW12237	Human secreted pro
426	6	1.9	125	22	AAW06825	Mitochondrial acyl
427	6	1.9	127	21	AAW06197	Human ORFX protein
428	6	1.9	129	22	AAW09463	Corynebacterium gl
429	6	1.9	129	23	AAW79711	Arabidopsis thalia
430	6	1.9	130	22	AAW3423	Human contig poly
431	6	1.9	131	21	AAW7933	lipid bioch thesis
432	6	1.9	131	22	AAW7933	P. patens lipid met
433	6	1.9	133	22	AAW80853	Human LPA-3 D2 reg
434	6	1.9	133	22	AAW82324	Proionbacterium
435	6	1.9	134	13	AAW55683	Human polypeptide
436	6	1.9	134	23	AAW88916	Human LPA-3-delta-
437	6	1.9	134	23	AAW88916	Actinomadura flexu
438	6	1.9	135	17	AAW94881	Proionbacterium
439	6	1.9	135	22	AAW63617	Novel human digno
440	6	1.9	137	22	ABG24903	Human ORFX protein
441	6	1.9	137	23	ABP06625	Arabidopsis thalia
442	6	1.9	140	21	AAW33784	Arabidopsis thalia
443	6	1.9	140	21	AAW33784	Drosophila melanog
444	6	1.9	140	21	AAW33784	C. glutamicum meta
445	6	1.9	142	22	AAW67122	Corynebacterium gl
446	6	1.9	142	22	AAW71893	Lactococcus lactis
447	6	1.9	142	22	AAW79665	
448	6	1.9	142	23	AAW53493	
449	6	1.9	144	21	AAW6306	Arabidopsis thalia
450	6	1.9	144	21	AAW68999	Hypervariable regi
451	6	1.9	144	23	ABP38054	Staphylococcus epi
452	6	1.9	144	23	ABW48853	Listeria monocytog
453	6	1.9	145	21	AAW33999	Arabidopsis thalia
454	6	1.9	145	21	ABP05983	Human ORFX protein
455	6	1.9	147	15	AAW60799	Rape abscission/de
456	6	1.9	148	21	AAW35834	Arabidopsis thalia
457	6	1.9	148	21	AAW35834	Arabidopsis thalia
458	6	1.9	148	21	AAW37201	Arabidopsis thalia
459	6	1.9	148	21	AAW38547	Arabidopsis thalia
460	6	1.9	148	21	AAW38547	Novel acid sequenc
461	6	1.9	148	21	AAW69006	Novel human digno
462	6	1.9	148	21	AAW69006	Arabidopsis thalia
463	6	1.9	150	21	AAW47796	Proionbacterium
464	6	1.9	150	22	AAW6351	Proionbacterium
465	6	1.9	150	22	AAW6351	Human AFP protein
466	6	1.9	151	21	AAW1363	Zea mays protein f
467	6	1.9	151	21	AAW1363	Drosophila melanog
468	6	1.9	151	22	AAW69434	Proionbacterium
469	6	1.9	151	22	AAW69434	C glutamicum prote
470	6	1.9	152	22	AAW62293	Novel human digno
471	6	1.9	152	22	AAW62293	C glutamicum prote
472	6	1.9	152	22	AAW62293	Arabidopsis pathog
473	6	1.9	154	21	AAW72380	Arabidopsis thalia
474	6	1.9	154	21	AAW72380	Drosophila melanog
475	6	1.9	154	21	AAW72380	Streptomyces negal
476	6	1.9	155	22	AAW91069	Novel human digno
477	6	1.9	156	22	AAW91069	Novel human digno
478	6	1.9	156	22	AAW91069	Novel human digno
479	6	1.9	159	21	AAW55981	Proionbacterium
480	6	1.9	160	22	AAW3406	Proionbacterium
481	6	1.9	161	22	AAW3406	Arabidopsis thalia
482	6	1.9	161	22	AAW3406	Streptococcus pneu
483	6	1.9	161	22	AAW3406	CER 89 protein seq
484	6	1.9	161	22	AAW3406	Novel human digno
485	6	1.9	162	22	ABG33475	Novel human digno
486	6	1.9	162	22	ABG33475	Arabidopsis thalia
487	6	1.9	162	22	ABG33475	Arabidopsis thalia
488	6	1.9	164	21	AAW3795	Novel human digno
489	6	1.9	164	21	AAW3795	Novel human digno
490	6	1.9	164	21	AAW3795	Novel human digno
491	6	1.9	166	22	ABG18026	Novel human digno
492	6	1.9	168	23	AAW87029	Adenovirus 5 fiber
493	6	1.9	169	21	AAW81561	Human N-acetylgluc
494	6	1.9	171	21	AAW81561	Human N-acetylgluc
495	6	1.9	172	21	AAW33645	Eucalyptus grandis
496	6	1.9	172	22	AAW33645	Peptide #6528 enco
497	6	1.9	172	22	AAW33645	Human bone marrow
498	6	1.9	172	22	AAW33645	Human bone marrow
499	6	1.9	172	22	AAW33645	Peptide #5947 enco
500	6	1.9	172	22	AAW33645	Human peptide enco
501	6	1.9	172	23	AAW33645	Human peptide enco
502	6	1.9	176	23	AAW33645	Arabidopsis thalia
503	6	1.9	177	22	AAW33645	Rattus norvegicus
504	6	1.9	177	22	AAW33645	Human immune/haema
505	6	1.9	177	22	AAW33645	Physcomitrella pat
506	6	1.9	177	22	AAW33645	Corynebacterium gl
507	6	1.9	181	22	AAW33645	Proionbacterium
508	6	1.9	181	22	AAW33645	Arabidopsis thalia
509	6	1.9	182	21	AAW33645	Arabidopsis thalia
510	6	1.9	182	21	AAW33645	Amino acid sequenc
511	6	1.9	185	21	AAW33645	Neisseria meningit
512	6	1.9	185	21	AAW33645	Neisseria meningit
513	6	1.9	185	21	AAW33645	Human polypeptide
514	6	1.9	187	22	AAW67725	Drosophila melanog
515	6	1.9	187	22	AAW67725	Novel human digno
516	6	1.9	187	23	ABG02156	Herbicidally activ
517	6	1.9	189	22	ABW82305	Novel human digno
518	6	1.9	191	22	ABG18348	Novel human digno
519	6	1.9	193	22	AAW12572	Novel signal trans
520	6	1.9	193	23	AAW12572	Human polypeptide
521	6	1.9	194	21	AAW36444	Arabidopsis thalia
522	6	1.9	194	22	AAW67524	Amino acid sequenc

522	6	1.9	196	22	AU20445	Human secreted pro
523	6	1.9	198	22	AAE02175	Mycobacterium xeno
524	6	1.9	198	22	AB878986	Lolium perenne lrp4
525	6	1.9	199	22	ABG33674	Novel human diagno
526	6	1.9	199	22	AAE02180	Mycobacterium xeno
527	6	1.9	199	22	AAE02182	Mycobacterium xeno
528	6	1.9	200	22	ABG07420	Novel human diagno
529	6	1.9	200	22	ABG07860	Novel human diagno
530	6	1.9	200	22	ABP29630	Streptococcus poly
531	6	1.9	202	22	AB859228	Drosophila melanog
532	6	1.9	202	22	AU57004	Propionibacterium
533	6	1.9	202	23	ABP03464	Human ORF proteinc
534	6	1.9	203	22	AAV37646	Amino acid sequenc
535	6	1.9	203	22	AAAG0062	C glutamincum prote
536	6	1.9	205	22	AU34710	E. coli cellular p
537	6	1.9	206	21	AAAG19275	Arabidopsis thalia
538	6	1.9	207	23	AA850648	Arabidopsis thalia
539	6	1.9	208	22	AU02481	Novel human secret
540	6	1.9	208	22	AAAG1792	C glutamincum prote
541	6	1.9	209	22	AAV28892	Lactobacillus brev
542	6	1.9	209	22	AU087471	Novel central nerv
543	6	1.9	209	22	AU09732	Human novel extrac
544	6	1.9	209	23	ABP47952	Human polypeptide
545	6	1.9	211	20	AAV25676	Eurogllyphus sp. al
546	6	1.9	211	20	AAV25677	Eurogllyphus sp. al
547	6	1.9	211	20	AA805952	Protein deduced fr
548	6	1.9	214	20	AAV29117	Amino acid sequenc
549	6	1.9	215	22	AB859408	Drosophila melanog
550	6	1.9	216	22	AU050081	Propionibacterium
551	6	1.9	218	21	AA808446	A human prostate s
552	6	1.9	218	21	AA808447	A human prostate s
553	6	1.9	220	19	AA869714	Streptomyces clavu
554	6	1.9	220	22	AB861843	Drosophila melanog
555	6	1.9	220	22	AA868323	N-acetyl transfera
556	6	1.9	221	22	AA896635	Putative P. abyssi
557	6	1.9	222	15	AA860313	Human mature LFA-3
558	6	1.9	223	22	AA806589	Human protein havi
559	6	1.9	223	22	AA898334	E. maynei Group 1
560	6	1.9	224	21	AA838018	Arabidopsis thalia
561	6	1.9	225	21	AA833478	Arabidopsis thalia
562	6	1.9	227	20	AA888925	Polypeptide fragme
563	6	1.9	227	22	AB850994	Human secreted pro
564	6	1.9	228	22	AU03676	Pseudomonas aerugi
565	6	1.9	230	22	AU04317	Propionibacterium
566	6	1.9	231	18	AA855544	H. pylori ORF 06ce
567	6	1.9	231	20	AAV17170	H. pylori outer me
568	6	1.9	231	23	AB853629	Lactococcus lactis
569	6	1.9	233	22	AU05694	G protein-coupled
570	6	1.9	234	18	AA811638	Human anti-RSV mon
571	6	1.9	234	23	AB847971	Listeria monocytog
572	6	1.9	236	15	AA845442	Sequence of the si
573	6	1.9	236	20	AA893968	Plasamid pBBP21 Dsb
574	6	1.9	236	21	AA846602	Amino acid sequenc
575	6	1.9	236	22	AA898307	Escherichia coli p
576	6	1.9	237	12	AA898881	E. coli growth and
577	6	1.9	237	12	AA841882	Human LFA-3 (CD58)
578	6	1.9	237	13	AA820804	Human LFA-3 antige
579	6	1.9	237	17	AA891432	Human LFA-3 antige
580	6	1.9	237	17	AA894020	Anti-erbB2 scFv.
581	6	1.9	237	18	AA816687	Human CD58 GPI. H
582	6	1.9	237	18	AA815185	Single-chain anti-
583	6	1.9	237	19	AA800441	Human LFA-3 antige
584	6	1.9	237	20	AA866189	Human LFA-3 antige
585	6	1.9	237	20	AA895440	A33/212 single-cha
586	6	1.9	237	21	AA896127	Human cell surface
587	6	1.9	237	21	AA891924	Arabidopsis thalia
588	6	1.9	237	21	AA897059	Amino acid sequenc
589	6	1.9	237	21	AA894836	Linked fusion prot
590	6	1.9	237	22	AU04095	Epsstein-Barr virus
591	6	1.9	237	22	AU02436	Human lymphocyte f
592	6	1.9	238	15	AA845443	Sequence of the si
593	6	1.9	238	21	AA821516	Arabidopsis thalia
594	6	1.9	238	22	AB868978	Drosophila melanog
595	6	1.9	238	23	AU092966	Arabidopsis transc
596	6	1.9	239	18	AA815186	Single-chain anti-
597	6	1.9	240	11	AA805572	PI-linked LFA-3 en
598	6	1.9	240	11	AA807604	Lymphocyte functio
599	6	1.9	240	13	AA827162	PI-linked LFA-3 us
600	6	1.9	240	14	AA834222	Amino acid sequenc
601	6	1.9	240	14	AA834372	Sequence encoded b
602	6	1.9	240	20	AA888095	Single-chain prote
603	6	1.9	240	21	AA837675	Bivalent antigen b
604	6	1.9	240	21	AA807931	A heterobivalent a
605	6	1.9	240	21	AA83134	PI-linked human tr
606	6	1.9	240	21	AA897250	4-4-20 VI region a
607	6	1.9	240	21	AA890920	Single chain antig
608	6	1.9	240	22	AA861158	Human PI-linked tr
609	6	1.9	240	23	AU076226	PI-linked lymphocy
610	6	1.9	241	14	AA837645	Sequence of a mult
611	6	1.9	241	20	AA895441	A33/218 single-cha
612	6	1.9	241	21	AA817852	Arabidopsis thalia
613	6	1.9	241	21	AA854837	Linked fusion prot
614	6	1.9	242	21	AA837570	Arabidopsis thalia
615	6	1.9	242	22	AB863185	Drosophila melanog
616	6	1.9	242	22	AB870551	Drosophila melanog
617	6	1.9	243	20	AA879551	Human M6p1 protei
618	6	1.9	243	22	AB860343	Drosophila melanog
619	6	1.9	243	22	AA890979	C glutamincum prote
620	6	1.9	244	20	AA888096	Single-chain prote
621	6	1.9	244	21	AA837676	Bivalent antigen b
622	6	1.9	244	21	AA807932	A heterobivalent a
623	6	1.9	244	21	AA857251	4-4-20 Vh region a
624	6	1.9	244	21	AA80921	Single chain antig
625	6	1.9	244	21	AA852290	Hamster acrosomal
626	6	1.9	245	20	AA897891	A33/218 single cha
627	6	1.9	247	21	AA833477	Arabidopsis thalia
628	6	1.9	248	14	AA837646	Sequence of a sing
629	6	1.9	249	20	AA897888	CC49/218 single ch
630	6	1.9	249	23	AU097275	Bacillus cereus ox
631	6	1.9	250	9	AA81507	Human LFA-3 expres
632	6	1.9	250	13	AA827161	Transmembrane LFA-
633	6	1.9	250	14	AA834221	Amino acid sequenc
634	6	1.9	250	14	AA834371	Sequence encoded b
635	6	1.9	250	14	AA837647	Sequence of a 4-4-
636	6	1.9	250	15	AA84271	Human LFA-3. Homo
637	6	1.9	250	17	AA804370	Human cell adhesio
638	6	1.9	250	20	AA888097	Single chain prote
639	6	1.9	250	21	AA827677	4-4-20/212 single
640	6	1.9	250	21	AA807933	A single-chain ant
641	6	1.9	250	21	AA817851	Arabidopsis thalia
642	6	1.9	250	21	AA838017	Arabidopsis thalia
643	6	1.9	250	21	AA831133	Human transmembran
644	6	1.9	250	21	AA857252	4-4-20/212 single-
645	6	1.9	250	21	AA880922	Single chain antig
646	6	1.9	250	22	AA867519	Amino acid sequenc
647	6	1.9	250	22	AA861157	Human transmembran
648	6	1.9	250	23	AU093167	Arabidopsis transc
649	6	1.9	250	23	AU076225	Lymphocyte functio
650	6	1.9	251	16	AA880192	Minor ampullate sp
651	6	1.9	251	18	AA820280	H. pylori cytoplasm
652	6	1.9	251	20	AA874015	Human prostate tum
653	6	1.9	251	22	AB803691	Novel human diagno
654	6	1.9	251	22	AA820361	Human bpg41608HDF
655	6	1.9	251	22	AA820361	Human hepatoma-der
656	6	1.9	251	23	ABP30062	Streptococcus poly
657	6	1.9	252	20	AAV00223	Enterococcus faeca
658	6	1.9	252	22	AB859632	Drosophila melanog
659	6	1.9	252	22	AA896231	Putative P. abyssi
660	6	1.9	252	23	ABP43442	E faecalis BF110 a
661	6	1.9	253	20	AA888098	Single chain prote
662	6	1.9	253	21	AA827678	4-4-20/212 single
663	6	1.9	253	21	AA807934	A single-chain ant
664	6	1.9	253	21	AA843712	Arabidopsis thalia
665	6	1.9	253	21	AA857253	4-4-20/212 single-
666	6	1.9	253	21	AA80923	Single chain antig
667	6	1.9	254	21	AA819874	Arabidopsis thalia

668	1.9	254	21	AAV69002	Amino acid sequenc	741	295	14	AAK41347	Nastertium xyloglu
669	1.9	254	23	ABP00351	Human ORFX protei	742	295	22	AAK4951	Shrimp white spot
670	1.9	255	21	AAK15125	Arabidopsis thalia	743	296	21	AAK13289	Arabidopsis thalia
671	1.9	255	23	AAU2763	Tomato DMR protei	744	296	21	AAK32305	Arabidopsis thalia
672	1.9	256	22	AAK2766	C glutamatum prote	745	296	21	AAK52739	Arabidopsis thalia
673	1.9	256	23	ABK2092	Herbicideally activ	746	296	21	AAK52739	Arabidopsis thalia
674	1.9	256	23	ABK2092	Herbicideally activ	747	297	21	AAK67526	Amino acid sequenc
675	1.9	257	20	AAK1524	Single chain-antic	748	297	22	AAK34835	Arabidopsis thalia
676	1.9	257	20	AAK1524	Single chain-antic	749	297	22	AAK34835	Arabidopsis thalia
677	1.9	257	21	AAK1514	Arabidopsis thalia	750	297	22	AAK34835	Arabidopsis thalia
678	1.9	257	21	AAK1514	Arabidopsis thalia	751	297	22	AAK34835	Arabidopsis thalia
679	1.9	257	22	AAK2902	C glutamatum prote	752	298	22	AAK17059	Arabidopsis thalia
680	1.9	257	22	AAK2902	C glutamatum prote	753	298	22	AAK17059	Arabidopsis thalia
681	1.9	257	23	AAU75157	Corynebacterium g1	754	299	21	AAK36551	Herbicideally activ
682	1.9	258	23	AAU10387	Trichoderma harzia	755	300	21	AAK36551	Herbicideally activ
683	1.9	258	23	AAU10387	Trichoderma harzia	756	300	21	AAK36551	Herbicideally activ
684	1.9	259	22	AAK47418	Marine MHRIT-AP62	757	301	21	AAK49028	Arabidopsis thalia
685	1.9	259	22	AAK47418	Marine MHRIT-AP62	758	301	21	AAK49028	Arabidopsis thalia
686	1.9	262	17	AAK97380	Epsa. Lactococcus	759	303	22	AAK88333	C. elegans mutant
687	1.9	262	18	AAK97380	CC49/212 SCA PLAP	760	303	22	AAK88333	C. elegans mutant
688	1.9	262	20	AAK97890	H. pylori cytoplas	761	306	21	AAK37568	Arabidopsis thalia
689	1.9	262	21	AAK97890	CC49/212 SCA PLAP	762	306	21	AAK37568	Arabidopsis thalia
690	1.9	262	22	AAK67518	Pseudomonas sp. WF	763	308	21	AAK51178	Arabidopsis thalia
691	1.9	263	12	AAK15055	On212 single chain	764	310	23	ABK60952	Novel human diagn
692	1.9	263	12	AAK15055	On212 single chain	765	311	22	ABK60952	Novel human diagn
693	1.9	264	17	AAK97381	Kabat consensus VK	766	312	23	ABK60952	Novel human diagn
694	1.9	264	19	AAK73049	Humantised A33 scAb	767	313	23	ABK60952	Novel human diagn
695	1.9	265	23	AAU75159	Novel floral meris	768	314	16	AAK87027	Novel human diagn
696	1.9	265	23	AAU75159	A33/218 sfv with c	769	314	16	AAK87027	Novel human diagn
697	1.9	267	23	AAK43823	Arabidopsis thalia	770	314	20	AAK35764	Novel human diagn
698	1.9	267	23	AAK43823	Arabidopsis thalia	771	314	20	AAK35764	Novel human diagn
699	1.9	268	19	AAK80319	Human DTHP polype	772	314	22	AAK36550	Novel human diagn
700	1.9	268	19	AAK80319	S. pneumoniae dexi	773	314	22	AAK36550	Novel human diagn
701	1.9	268	23	ABK60931	Drosophila melano	774	315	22	AAK36550	Novel human diagn
702	1.9	268	23	ABK60931	Novel floral meris	775	315	22	AAK36550	Novel human diagn
703	1.9	268	23	AAU87039	Adenovirus 5 recom	776	316	22	AAK36550	Novel human diagn
704	1.9	269	23	AAU87040	Adenovirus 5 recom	777	317	22	AAK36550	Novel human diagn
705	1.9	270	21	AAK1918	CC49/218 sfv #2 w1	778	318	22	AAK36550	Novel human diagn
706	1.9	270	21	AAK1918	Bacillus lentus su	779	319	20	AAK36550	Novel human diagn
707	1.9	271	21	AAV70713	Arabidopsis thalia	780	319	22	AAK36550	Novel human diagn
708	1.9	271	21	AAV70713	Rat p35 (p53) isofo	781	319	22	AAK36550	Novel human diagn
709	1.9	271	21	AAV68995	Amino acid sequenc	782	320	21	AAK36550	Novel human diagn
710	1.9	271	21	AAV68995	Amino acid sequenc	783	321	22	AAK36550	Novel human diagn
711	1.9	271	21	AAV68995	Amino acid sequenc	784	321	22	AAK36550	Novel human diagn
712	1.9	274	21	AAK15287	Novel human diagn	785	321	22	AAK36550	Novel human diagn
713	1.9	274	21	AAK15287	Novel human diagn	786	321	22	AAK36550	Novel human diagn
714	1.9	275	21	AAV28888	Hyperxanthinase regl	787	321	22	AAK36550	Novel human diagn
715	1.9	275	21	AAV28888	Lactobacillus brev	788	321	22	AAK36550	Novel human diagn
716	1.9	276	22	AAK96482	Putative P. abyssi	789	321	22	AAK36550	Novel human diagn
717	1.9	279	20	AAV39341	Staphylococcus aur	790	322	21	AAK36550	Novel human diagn
718	1.9	279	20	AAV39341	S. aureus spoc07 p	791	322	21	AAK36550	Novel human diagn
719	1.9	279	21	AAK15290	Arabidopsis thalia	792	323	23	AAK36550	Novel human diagn
720	1.9	279	21	AAK15290	Arabidopsis thalia	793	324	23	AAK36550	Novel human diagn
721	1.9	279	21	AAK15290	Arabidopsis thalia	794	324	23	AAK36550	Novel human diagn
722	1.9	279	21	AAK15290	Arabidopsis thalia	795	325	23	AAK36550	Novel human diagn
723	1.9	279	21	AAK15290	Arabidopsis thalia	796	325	23	AAK36550	Novel human diagn
724	1.9	280	21	AAK15290	Human polypeptide	797	326	21	AAK36550	Novel human diagn
725	1.9	281	23	ABK2102	Herbicideally activ	798	327	22	AAK36550	Novel human diagn
726	1.9	282	23	ABK2102	Herbicideally activ	799	327	22	AAK36550	Novel human diagn
727	1.9	283	23	AAU75160	Amino acid sequenc	800	328	22	AAK36550	Novel human diagn
728	1.9	283	23	AAU75160	S. erythraea pfuJ7	801	328	22	AAK36550	Novel human diagn
729	1.9	284	23	ABK4441	Kabat consensus VK	802	328	22	AAK36550	Novel human diagn
730	1.9	285	22	AAU43138	Enterococcus faeca	803	329	23	AAK36550	Novel human diagn
731	1.9	285	22	AAU43138	E faecalis EPI10 p	804	329	23	AAK36550	Novel human diagn
732	1.9	286	22	AAU47052	R. prowazekii Spap	805	329	23	AAK36550	Novel human diagn
733	1.9	287	23	ABK39757	Propionibacterium	806	330	23	AAK36550	Novel human diagn
734	1.9	287	23	ABK39757	Propionibacterium	807	330	23	AAK36550	Novel human diagn
735	1.9	287	23	ABK39757	Propionibacterium	808	331	20	AAK36550	Novel human diagn
736	1.9	290	21	AAK15179	Staphylococcus epi	809	331	20	AAK36550	Novel human diagn
737	1.9	290	21	AAK15179	Amino acid sequenc	810	331	21	AAK36550	Novel human diagn
738	1.9	292	23	ABK27029	Rice M10 homologue	811	331	21	AAK36550	Novel human diagn
739	1.9	293	22	ABK15628	Arabidopsis thalia	812	331	22	AAK36550	Novel human diagn
740	1.9	295	11	AAK07605	Novel human diagn	813	331	23	AAK36550	Novel human diagn

814	6	1.9	333	20	AAW66019	Murine homeobox HO
815	6	1.9	333	21	AAV70111	Anti-Herz neu singl
816	6	1.9	334	18	AAW40803	Alternate sigma fa
817	6	1.9	334	18	AAW38360	Pseudomonas fluores
818	6	1.9	334	18	AAW30203	Rpos coding sequen
819	6	1.9	334	19	AAW31748	Pseudomonas altern
820	6	1.9	336	22	ABB52799	Escherichia coli p
821	6	1.9	336	22	ABP40452	Staphylococcus epi
822	6	1.9	337	21	AAV74311	Neisseria gonorrhoe
823	6	1.9	338	22	ABB65069	Drosophila melanog
824	6	1.9	338	22	ABB77701	Amino acid sequenc
825	6	1.9	339	22	AAW35358	Human polypeptide,
826	6	1.9	339	22	AAW30497	C glutaminc prote
827	6	1.9	339	22	AAW32513	Corynebacterium gl
828	6	1.9	339	22	AAW67555	Corynebacterium gl
829	6	1.9	339	22	AAW67556	C. elegans mutant
830	6	1.9	339	23	ABB78674	Arabidopsis thalia
831	6	1.9	340	21	AAW19273	Herbicidally activ
832	6	1.9	340	23	ABB92701	Enterococcus faeca
833	6	1.9	341	22	AAW5076	Human Sec2 protein
834	6	1.9	343	19	AAW69332	Bacteriophage T7 c
835	6	1.9	343	21	AAW5918	Arabidopsis thalia
836	6	1.9	343	21	AAW32303	Escherichia coli p
837	6	1.9	343	22	ABB52553	Human Sec2 catalyt
838	6	1.9	344	21	AAW36106	Human cancer assoc
839	6	1.9	344	21	AAW3522	Amino acid sequenc
840	6	1.9	344	21	AAW18800	Human secreted pro
841	6	1.9	344	22	ABB51281	Human gene 1 encod
842	6	1.9	345	22	AAW2427	Stem region of A/O
843	6	1.9	348	15	AAW3589	Human prostate tum
844	6	1.9	352	20	AAW74074	Adenovirus 5 fiber
845	6	1.9	352	23	AAW87031	Rat hepatoma H35 c
846	6	1.9	353	21	AAW36105	Rat hepatoma H35 c
847	6	1.9	353	21	AAW16623	Novel human enzyme
848	6	1.9	354	22	AAW23056	Novel human enzyme
849	6	1.9	354	22	AAW23622	Cattle IgG transpo
850	6	1.9	354	22	AAW82604	Sheep IgG transpor
851	6	1.9	354	22	AAW82606	C glutaminc prote
852	6	1.9	354	22	AAW91346	Corynebacterium gl
853	6	1.9	354	22	AAW79499	Listeria monocytog
854	6	1.9	354	23	AAW49193	Polyamide hydrolas
855	6	1.9	355	18	AAW29999	Polyamide hydrolas
856	6	1.9	355	18	AAW36011	Human polypeptide
857	6	1.9	355	22	AAW40927	Staphylococcus aur
858	6	1.9	356	22	AAW37021	Staphylococcus aur
859	6	1.9	356	22	AAW37260	Human gene 22-enco
860	6	1.9	357	22	AAW71391	Arabidopsis thalia
861	6	1.9	358	21	AAW20364	Human polypeptide
862	6	1.9	358	23	AAW80371	C glutaminc prote
863	6	1.9	360	22	AAW91214	E. coli esterase p
864	6	1.9	363	19	AAW66101	Burkholderia cepac
865	6	1.9	363	19	AAW61276	Burkholderia cepac
866	6	1.9	363	22	AAW73542	C. glutaminc SRT
867	6	1.9	364	22	AAW78974	Parathion hydrolas
868	6	1.9	365	11	AAW05573	GDP-Fuc-beta-D-gal
869	6	1.9	365	12	AAW13751	A glycosyltransfer
870	6	1.9	365	15	AAW45936	GDP-L-fucose-beta
871	6	1.9	365	16	AAW80154	2-Alpha-fucosyltra
872	6	1.9	365	16	AAW70422	Human H-transferr
873	6	1.9	365	16	AAW0572	Human H-transferr
874	6	1.9	365	17	AAW90572	Human alpha-1,2-fu
875	6	1.9	365	18	AAW3805	Human alpha-1,2-fu
876	6	1.9	365	18	AAW13640	Human alpha(1,2)-f
877	6	1.9	365	19	AAW53102	Pig H transferrase
878	6	1.9	365	20	AAW97356	Swine alpha(1,2) f
879	6	1.9	365	20	AAW30630	Swine alpha-1,2-fu
880	6	1.9	365	21	AAW97279	Human H-transferr
881	6	1.9	365	21	AAW79302	Pig alpha-1-2 fuc
882	6	1.9	365	22	AAW42255	Propionibacterium
883	6	1.9	365	22	AAW7995	Swine alpha (1,2)
884	6	1.9	365	23	AAW3695	Herbicidally activ
885	6	1.9	367	22	AAW51280	Human secreted pro
886	6	1.9	367	23	AAW87035	Adenovirus 5 fiber
887	6	1.9	368	21	AAW94405	Human ACMA4/IgG1-F
888	6	1.9	369	17	AAW02253	Human transcriptio
889	6	1.9	370	21	AAW62214	Arabidopsis thalia
890	6	1.9	371	23	AAW74617	Oestrogen-regulate
891	6	1.9	371	23	AAW71224	Novel signal trans
892	6	1.9	373	22	AAW37720	Human protein sequ
893	6	1.9	376	22	AAW61405	Novel human diagn
894	6	1.9	378	21	AAW6409	Drosophila melanog
895	6	1.9	378	21	AAW30906	Drosophila odorant
896	6	1.9	378	21	AAW62213	Arabidopsis thalia
897	6	1.9	378	22	AAW66092	Drosophila melanog
898	6	1.9	378	22	AAW32975	C glutaminc prote
899	6	1.9	379	21	AAW26408	Drosophila melanog
900	6	1.9	379	21	AAW30907	Drosophila odorant
901	6	1.9	379	22	AAW69965	Drosophila melanog
902	6	1.9	379	22	ABP27052	Streptococcus poly
903	6	1.9	379	23	ABP07498	Human lipid metabo
904	6	1.9	380	16	AAW73999	Rhodococcus rhodoc
905	6	1.9	380	21	AAW36104	Rat hepatoma H35 c
906	6	1.9	380	23	AAW02093	Protein of nitrica
907	6	1.9	381	23	AAW16622	Rat hepatoma H35 c
908	6	1.9	381	22	AAW6576	Cysteine protease
909	6	1.9	382	22	AAW63575	Drosophila melanog
910	6	1.9	382	22	AAW91079	C glutaminc prote
911	6	1.9	383	23	AAW51460	Human CDV-1-42. H
912	6	1.9	385	19	AAW62644	Flea serine protea
913	6	1.9	385	21	AAW62644	Arabidopsis thalia
914	6	1.9	385	21	AAW36212	Arabidopsis thalia
915	6	1.9	385	21	AAW39363	Arabidopsis thalia
916	6	1.9	385	21	AAW60555	Arabidopsis thalia
917	6	1.9	386	21	AAW68459	PERSV isolate 14/9
918	6	1.9	386	23	AAW78648	Caenorhabditis ele
919	6	1.9	386	23	AAW78660	C. elegans mutant
920	6	1.9	386	23	AAW78661	C. elegans mutant
921	6	1.9	386	23	AAW78664	C. elegans mutant
922	6	1.9	386	23	AAW78665	C. elegans mutant
923	6	1.9	386	23	AAW78667	C. elegans mutant
924	6	1.9	386	23	AAW78668	C. elegans mutant
925	6	1.9	386	23	AAW78670	C. elegans mutant
926	6	1.9	386	23	AAW78675	C. elegans mutant
927	6	1.9	388	22	AAW66813	Drosophila melanog
928	6	1.9	388	23	AAW81727	Full length human
929	6	1.9	389	21	AAW59209	B. halodurans poly
930	6	1.9	389	23	AAW28308	Streptococcus coly
931	6	1.9	390	21	AAW1640	Human ORX ORF1404
932	6	1.9	391	23	AAW28307	Streptococcus poly
933	6	1.9	392	22	AAW93576	Human polypeptide,
934	6	1.9	393	19	AAW68458	PERSV isolate 4/96
935	6	1.9	393	20	AAW9471	Maize uroporphyrin
936	6	1.9	394	21	AAW29720	Arabidopsis thalia
937	6	1.9	394	22	AAW47605	Uroporphyrinogen d
938	6	1.9	394	22	AAW20018	Arabidopsis uropor
939	6	1.9	395	23	AAW51424	Human MDR SEQ ID
940	6	1.9	397	19	AAW62633	Flea serine protea
941	6	1.9	397	19	AAW62628	Flea serine protea
942	6	1.9	398	19	AAW80405	A secreted protein
943	6	1.9	398	20	AAW29592	Human MGPI protei
944	6	1.9	398	20	AAW13358	Amino acid sequenc
945	6	1.9	398	21	AAW94403	Human ACM cellula
946	6	1.9	398	21	AAW62988	Amino acid sequenc
947	6	1.9	398	21	AAW45095	Human LDCAM bindin
948	6	1.9	398	22	AAW12345	Human PRO258 polyp
949	6	1.9	398	22	AAW80226	Human PRO258 prote
950	6	1.9	398	22	AAW171948	Human ion channel
951	6	1.9	398	22	AAW53083	Human angioogenesis
952	6	1.9	398	23	AAW51823	Human polypeptide
953	6	1.9	398	23	AAW85444	Human angioogenesis
954	6	1.9	398	23	AAW84838	Human PRO258 polyp
955	6	1.9	398	23	AAW17448	Human sodium ion c
956	6	1.9	399	22	AAW64560	Drosophila melanog
957	6	1.9	400	21	AAW14139	Bordetella pertuss
958	6	1.9	400	22	AAW86933	C. glutaminc ATCC
959	6	1.9	402	21	AAW60554	Arabidopsis thalia

960	6	1.9	402	22	ABR64529	Drosophila melanog
961	6	1.9	403	21	ABG20948	Arabidopsis thaliana
962	6	1.9	403	21	AAE59362	Arabidopsis thaliana
963	6	1.9	403	22	AAE69362	C. glutamicum ATCC
964	6	1.9	405	22	ABG28543	Novel human diseno
965	6	1.9	407	21	AAV78271	Streptomyces coelic
966	6	1.9	407	21	AAV78823	Amino acid sequenc
967	6	1.9	407	22	AAE31679	C. glutamicum prote
968	6	1.9	407	23	AAE31583	Streptomyces fradi
969	6	1.9	408	23	ABE35521	Lactococcus lactis
970	6	1.9	409	22	AAE39391	Lactococcus lactis
971	6	1.9	410	21	AAV32307	Rice acid triacylg
972	6	1.9	411	18	AAW01558	TYPI protein. Can
973	6	1.9	411	21	AAV15114	Candida cell-cycle
974	6	1.9	412	22	ABW71197	Drosophila melanog
975	6	1.9	412	22	AAE78973	C. glutamicum SRT
976	6	1.9	413	22	AAE00867	Human brain immuno
977	6	1.9	415	22	AAU17462	Novel signal trans
978	6	1.9	415	23	AAE66148	45 kDa protein of
979	6	1.9	416	19	AAE60836	Human LAMP family
980	6	1.9	416	20	AAE55927	Human STRK2 protei
981	6	1.9	416	20	AAV22605	Human TSC403 prote
982	6	1.9	416	20	AAV22648	A murine signal tra
983	6	1.9	416	20	AAV22651	Murine 5e. new poly
984	6	1.9	416	20	AAV21673	Human Ste20-like s
985	6	1.9	416	21	AAV82274	Mouse protein sequ
986	6	1.9	416	22	AAV82277	Human polypeptide
987	6	1.9	416	22	AAW40348	Human h2252. beta. rec
988	6	1.9	416	22	AAE6607	Human TGR-beta. rec
989	6	1.9	417	21	AAE01209	Corn putative lecti
990	6	1.9	417	22	AAE71349	Human alternative
991	6	1.9	417	22	AAE71349	Human WTL variant
992	6	1.9	418	15	AAE60135	Comamonas testocoe
993	6	1.9	418	15	AAE60135	Linoleic-acid-dessa
994	6	1.9	420	18	AAE60501	M. gordoniae gyrase
995	6	1.9	420	18	AAW15076	M. flavescentis gyra
996	6	1.9	421	18	AAW15075	M. flavescentis gyra
997	6	1.9	421	18	AAW15075	Drosophila melanog
998	6	1.9	422	21	AAE64654	Arabidopsis thaliana
999	6	1.9	422	21	AAE52741	Human polypeptide,
1000	6	1.9	422	22	AAE93395	

ALIGNMENTS

RESULT 1
AAB20106 standard; Protein; 322 AA.

XX	AC	AAB20106;	
XX	DT	23-APR-2001 (first entry)	
XX	DE	Moraxella catarrhalis BASB110 protein.	
XX	XX	BASB110; infection; otitis media; pneumonia; therapy; diagnosis;	
XX	KM	antibacterial; antimicrobial; vaccine.	
XX	OS	Moraxella catarrhalis.	
XX	XX	Key	Location/Qualifiers
XX	FT	Misc-difference 104	
XX	FT	/note="Gly in translation of BASB110 PCR product"	
XX	XX	WO200100838-A1.	
XX	XX	04-JAN-2001.	
XX	PD	23-JUN-2000; 2000WO-EP05854.	
XX	PF	23-JUN-2000; 2000WO-EP05854.	
XX	PR	25-JUN-1999; 99GB-0015031.	
XX	XX		

PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX	Thomard J;
XX	WPI: 2001-112459/12.
XX	N-FSDS; AAF30046.
XX	Novel BASB110 polypeptides of Moraxella catarrhalis, useful as a
XX	vaccine for treating Moraxella catarrhalis infections -
XX	Claim 1; Page 82-83; 88pp; English.
XX	The present sequence is that of BASB110 protein from Moraxella
XX	catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
XX	media in children and pneumonia in adults. The sequence is a
XX	translation of the BASB110 gene coding region (see AAF30046). It
XX	differs by 1 residue from the predicted polypeptide sequence (see
XX	AAE20107) of a BASB110 PCR product, having Ser rather than Gly at
XX	position 104. The invention provides BASB110 polypeptides, and
XX	polynucleotides encoding them, as well as expression vectors, host
XX	cells and methods for producing BASB110 polypeptides using
XX	recombinant methods. Also claimed is a vaccine composition
XX	comprising a BASB110 polypeptide, an immunogenic fragment of a
XX	BASB110 polypeptide, or a polypeptide having at least 85% amino
XX	acid sequence identity to BASB110, or comprising a polynucleotide
XX	encoding such a polypeptide. A claimed method of diagnosing a
XX	Moraxella infection involves identifying a BASB110 polypeptide or
XX	antibody. A claimed therapeutic composition useful in treating
XX	humans with M. catarrhalis infection comprises at least 1 antibody
XX	directed against a BASB110 polypeptide. BASB110 polypeptides also
XX	have utility in raising specific antibodies, and in screening for
XX	antibacterial drugs.
XX	Sequence 322 AA;
XX	Query Match 100.0%; Score 322; DB 22; Length 322;
XX	Best Local Similarity 100.0%; Pred. No. 1.2e-310; Indels 0; Gaps 0;
XX	Matches 322; Conservative 0; Mismatches 0; Indels 0;
QY	1 MVTITAINSONQKPKRLGLIFGVITTCILAGCSKPTYNSTSGSHRTSGSGGLAIGS 60
DB	1 MVTITAINSONQKPKRLGLIFGVITTCILAGCSKPTYNSTSGSHRTSGSGGLAIGS 60
QY	61 QVITDSQGVNRYOVKQDVTYSKIAORGLNWRREIGHINNINSSTYITYGMLTMSGDL 120
DB	61 QVITDSQGVNRYOVKQDVTYSKIAORGLNWRREIGHINNINSSTYITYGMLTMSGDL 120
QY	121 KYRERSISGGVNTAHTSPFVAVOSSRPVQOHPAVOKETPPVYVYKKETPPVVOQAP 180
DB	121 KYRERSISGGVNTAHTSPFVAVOSSRPVQOHPAVOKETPPVYVYKKETPPVVOQAP 180
QY	181 VAPPTTEAPFATGSGGVQFRYPVATVPVRRFGTAVAGSTVYNSGMWFSGRGDILIN 240
DB	181 VAPPTTEAPFATGSGGVQFRYPVATVPVRRFGTAVAGSTVYNSGMWFSGRGDILIN 240
QY	241 ASNAGTVIQADHNDGASIVIOHTNGFVSYTHIKDAQVKTGDPVRTGQRTASMKNQPSG 300
DB	241 ASNAGTVIQADHNDGASIVIOHTNGFVSYTHIKDAQVKTGDPVRTGQRTASMKNQPSG 300
QY	301 AALPEFRISRNQGVYVDPPLTVLK 322
DB	301 AALPEFRISRNQGVYVDPPLTVLK 322
XX	RESULT 2
XX	AAB20107 standard; Protein; 322 AA.
XX	XX AAB20107;
XX	XX 23-APR-2001 (first entry)
XX	XX Moraxella catarrhalis BASB110 protein.

```

XX KW BASB110; infection; otitis media; pneumonia; therapy; diagnosis;
XX KW antibacterial; antimicrobial; vaccine.
XX OS Moraxella catarrhalis.
FH Key Location/Qualifiers
FT Misc-difference 104
FT /note="Ser in translation of BASB110 gene"
XX PN MO200100838-A1.
XX PD 04-JAN-2001.
XX PP 23-JUN-2000; 2000WO-EP05854.
XX PR 25-JUN-1999; 99GB-0015031.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Thonard J;
XX DR WPI: 2001-112459/12.
XX DR N-PSDB; AAF30047.
XX PT Novel BASB110 polypeptides of Moraxella catarrhalis, useful as a
XX PS vaccine for treating Moraxella catarrhalis infections -
XX PS Claim 1; Page 83-84; 88pp; English.
XX CC The present sequence is that of BASB110 protein from Moraxella
XX CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
XX CC media in children and pneumonia in adults. The sequence is a
XX CC translation of a BASB110 PCR product (see AAF30047). It differs by
XX CC 1 residue from the predicted polypeptide sequence (see AAB20106) of
XX CC the BASB110 gene product, having Gly rather than Ser at position
XX CC 104. The invention provides BASB110 polypeptides, and
XX CC polynucleotides encoding them, as well as expression vectors, host
XX CC cells and methods for producing BASB110 polypeptides using
XX CC recombinant methods. Also claimed is a vaccine composition
XX CC comprising a BASB110 polypeptide, an immunogenic fragment of a
XX CC BASB110 polypeptide, or a polypeptide having at least 85% amino
XX CC acid sequence identity to BASB110, or comprising a polynucleotide
XX CC encoding such a polypeptide. A claimed method of diagnosing a
XX CC Moraxella infection involves identifying a BASB110 polypeptide or
XX CC antibody. A claimed therapeutic composition useful in treating
XX CC humans with M. catarrhalis infection comprises at least 1 antibody
XX CC directed against a BASB110 polypeptide. BASB110 polypeptides also
XX CC have utility in raising specific antibodies, and in screening for
XX CC antibacterial drugs.
XX SQ Sequence 322 AA;
Query Match 68.6%; Score 221; DB 22; Length 322;
Best Local Similarity 99.7%; Pred. No. 1.7e-210;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTVTIATNSONOKIKRLGLFVITTCIAGCASKPTYNSTSGSGSRISGSGLAIGS 60
DB 1 MTVTIATNSONOKIKRLGLFVITTCIAGCASKPTYNSTSGSGSRISGSGLAIGS 60
QY 61 QVITDSOGVPRVYVKGDTVSKIAQRXGLNMRREIGHINNNSYTIYTGOM/LTMSGD 120
DB 61 QVITDSOGVPRVYVKGDTVSKIAQRXGLNMRREIGHINNNSYTIYTGOM/LTMSGD 120
QY 121 KVRERSISISSGVNVAHTSPVAVOGSSRPVQHPAVQKPTPPVVVVKKPTPPVVOQAP 180
DB 121 KVRERSISISSGVNVAHTSPVAVOGSSRPVQHPAVQKPTPPVVVVKKPTPPVVOQAP 180
QY 121 KVRERSISISSGVNVAHTSPVAVOGSSRPVQHPAVQKPTPPVVVVKKPTPPVVOQAP 180
DB 121 KVRERSISISSGVNVAHTSPVAVOGSSRPVQHPAVQKPTPPVVVVKKPTPPVVOQAP 180
QY 181 VAPVTEAPFATGSSGVWQRFYVGATNPVVRRTAGTAVASTYNSNGMFFSGRDGLIN 240
DB 181 VAPVTEAPFATGSSGVWQRFYVGATNPVVRRTAGTAVASTYNSNGMFFSGRDGLIN 240

```

```

QY 241 ASNAGTVIQADHNDGASIVIQHTNGFVSSYIHIKDAQVKTGDTVRTGQRIASMKNOPSG 300
DB 241 ASNAGTVIQADHNDGASIVIQHTNGFVSSYIHIKDAQVKTGDTVRTGQRIASMKNOPSG 300
QY 301 AALFEFRISRNGVYVDPPLTVLK 322
DB 301 AALFEFRISRNGVYVDPPLTVLK 322
RESULT 3
ABB61691
ID ABB61691 standard; Protein; 968 AA.
XX AC ABB61691;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 11865.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI: 2001-656860/75.
XX DR N-PSDB; ABL05794.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 11865; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins.
XX CC Sequences (ABL01840-ABL16175) and the encoded proteins.
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 968 AA;
Query Match 2.8%; Score 9; DB 22; Length 968;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 217 ATVAGSTVT 225
DB 877 ATVAGSTVT 885
RESULT 4
AAG90076
ID AAG90076 standard; Protein; 223 AA.
XX

```


XX 02-MAR-2001; 2001WO-US06962.
 PF
 XX
 PR 03-MAR-2000; 2000US-187209P.
 PR 03-AUG-2000; 2000US-222898P.
 XX
 PR (UTAH) UNIV UTAH RES FOUND.
 PR
 XX Placek L, White S, Fu Y, Skradski S;
 PR WPI; 2001-589903/66.
 XX
 PR Monogenic Audiogenic Seizure-Susceptible-1 (mass1) genes, useful for
 PT producing animal models of audiogenic seizures -
 XX
 XX Disclosure; Fig 7; 79pp; English.
 XX
 XX The present invention relates to nucleic acid molecules designated
 CC monogenic audiogenic seizure-susceptible-1 (mass1) genes. The nucleic
 CC acid molecule may be used via recombinant DNA methodologies in the
 CC production of transgenic animal (especially mouse) models for studying
 CC genetic abnormalities related to mass1 which result in seizure
 CC susceptible phenotypes (mass1 is audiogenic seizures). The present
 CC sequence is mouse monogenic audiogenic seizure-susceptible (mass1)
 CC protein conserved amino acid repeat motif.
 CC Note: The present sequence (SEQ ID NO:16) is incorrectly referred to as
 CC SEQ ID NO:15 in the sequence listing of the specification.
 XX
 XX Sequence 35 AA;
 SQ
 Query Match 2.2%; Score 7; DB 22; Length 35;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 287 TGGRIAS 293
 Db 7 TGGRIAS 13
 RESULT 9
 ABP04087 standard; Protein; 54 AA.
 XX
 AC ABP04087;
 XX
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX protein sequence SEQ ID NO:8156.
 XX
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumor; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hyperthyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.
 KW
 OS Homo sapiens.
 XX
 PN WO200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US10836.
 XX
 PR 30-MAY-2000; 2000US-206132P.
 PR 29-AUG-2000; 2000US-228716P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach MD;
 XX

DR WPI; 2002-106308/14.
 DR N-PSDB; AEN19839.
 XX
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -
 XX
 XX Disclosure; SEQ ID 8156; 1037pp; English.
 XX
 XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where x is 1-11491 (see Table 1
 CC in the specification). AEN15762 to AEN27257 encode the human ORFX
 CC proteins given in ABP0010 to ABP1500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic
 CC transplantation, cardiovascular diseases, disorders related to organ
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 54 AA;
 SQ
 Query Match 2.2%; Score 7; DB 23; Length 54;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 178 PAPVAPP 184
 Db 4 PAPVAPP 10
 RESULT 10
 AAM83197 standard; Protein; 66 AA.
 XX
 AC AAM83197;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen SEQ ID NO:10790.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis.
 KW
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225265.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-483426/52.

N-PSDB; AAK55978.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
metastasis -

Claim 11; SEQ ID NO 10790; 3071pp + Sequence Listing; English.

CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
amino acid sequences given in AAK62170 to AAK91921. (I) have cytosolic

CC activity, and can be used in gene therapy and vaccine production. (1)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (1) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (1) by expressing inactive proteins or to
 CC supplement the patient's own production of (1). Additionally, (1)
 CC polynucleotides may be used to produce the secreted (1), by inserting the
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (1) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK67694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM62169
 CC represent sequences used in the exemplification of the present invention.
 CC
 CC Sequence 66 AA:

Query Match 2.2% Score 7; DB 22; Length 66;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

166 KKPTPTP 172
 11 KKPTPTP 17

RESULT 11
 ID AAY41355 standard; Protein; 67 AA.

AC AAY41355;
 DT 02-DEC-1999 (first entry)

XX Human secreted protein encoded by gene 48 clone HMAAP70.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 XX developmental abnormality; foetal deficiency; blood; allergy; renal;
 XX immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 XX inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 XX osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 XX endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX MO9947540-A1.

XX 23-SEP-1999.

XX 18-MAR-1999; 99WO-US05804.

XX 19-MAR-1998; 98US-0078563.
 XX 19-MAR-1998; 98US-0078566.
 XX 19-MAR-1998; 98US-0078573.
 XX 19-MAR-1998; 98US-0078574.
 XX 19-MAR-1998; 98US-0078576.
 XX 19-MAR-1998; 98US-0078577.
 XX 19-MAR-1998; 98US-0078578.
 XX 19-MAR-1998; 98US-0078579.
 XX 19-MAR-1998; 98US-0078581.
 XX 01-APR-1998; 98US-0080312.
 XX 01-APR-1998; 98US-0080313.
 XX 01-APR-1998; 98US-0080314.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
 XX Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW;
 XX Olsen HS, Shi Y, Moore PA;

XX

DR WPI; 1999-562050/47.
 DR N-PSDB; AA224858.

PT New isolated human genes, useful for diagnosis and treatment of e.g.
 PT cancers, neurological disorders, immune diseases, inflammation or blood
 PT disorders -

PS Claim 11; Page 387; 484p; English.

CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AA224802) for increasing the stability of the fused
 CC protein as compared to the human protein only.

CC The invention relates to 95 novel genes and their fragments (nucleic
 CC acid sequences: AA224811-224907; amino acid sequences AAY41308-Y41404)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC polypeptides in a sample or by determining the amount of the new
 CC the new polynucleotides. Specific uses are described for each of the 95
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AA224811 for described uses).

SO Sequence 67 AA;

Query Match 2.2% Score 7; DB 20; Length 67;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LGLRNV 24
 11 LGLRNV 17

RESULT 12
 ID AAU61587 standard; Protein; 72 AA.

AC AAU61587;
 DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #22483.

XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
 XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 XX dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX MO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.
 XX 02-JUN-2000; 2000US-200841P.
 XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 XX L'maisonmeuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59618.

PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -

XX Example 1; SEQ ID No 22782; 1069pp; English.
PS
XX Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
SQ
XX Sequence 72 AA;
Query Match 2.2%; Score 7; DB 22; Length 72;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 178 PAPVAPP 184
|||
Db 28 PAPVAPP 34
RESULT 13
AAU50707
ID AAU50707 standard; Protein; 80 AA.
XX
AC AAU50707;
XX
DT 27-FEB-2002 (first entry)
XX
DE *Propionibacterium acnes* immunogenic protein #11603.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS *Propionibacterium acnes*.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PE 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhattacha A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59549.
XX
PT *Propionibacterium acnes* polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -

XX Example 1; SEQ ID No 11902; 1069pp; English.
PS
XX Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
SQ
XX Sequence 80 AA;
Query Match 2.2%; Score 7; DB 22; Length 80;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 SGSGSHR 49
|||
Db 43 SGSGSHR 49
RESULT 14
AAU96225
ID AAU96225 standard; Protein; 90 AA.
XX
AC AAU96225;
XX
DT 11-SEP-2000 (first entry)
XX
DE *Arabidopsis* LEC1 consensus region.
XX
KW Leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
KW selectable marker; transgenic plant; transgenic seed; HAP3.
XX
OS *Arabidopsis* sp.
XX
PN WO200028058-A2.
XX
PD 18-MAY-2000.
XX
PE 09-NOV-1999; 99WO-US26514.
XX
PR 09-NOV-1998; 98US-0107643.
PR 10-NOV-1998; 98US-0107810.
XX
PA (PION-) PIONEER HI-BRED INT INC.
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
PI Sun X, Hoerster GU, Gregory CA, Nadimpalli R;
XX
DR WPI; 2000-376568/32.
XX
PT New HAP3-type CCAAT-box binding transcriptional activators,
PT particularly leafy cotyledon 1 transcriptional activator, useful for
PT inducing somatic embryogenesis or apomixis in a plant cell -
PS Example 6; Fig 1, 94pp, English.

The present sequence is the leafy cotyledon 1 transcriptional activator, LEC1 consensus sequence from Arabidopsis. This sequence is a HAP3-type CCAAT-box binding protein. LEC1 expression initiates the formation of embryo-like structures and improves growth and recovery of transformants. When apomixis occurs, i.e. the replacement of sexual reproduction by asexual reproduction, LEC1 expression in the nucellus integument, or cell specific expression in the megaspore mother cell would trigger embryo formation from maternal tissues only. This results in the production of seeds identical to the parent. In addition, LEC1 transgenic high yielding seeds could be developed. In addition, LEC1 plant could be used for positive selection of a transformed cell (transgenic plant) for increasing transformation efficiency and for increasing recovery of regenerated plants. The present sequence was used in the identification of the plant LEC1 consensus sequence (AY96223).

Sequence 90 AA;

Query Match 2.2%; Score 7; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 314 YVDPPLTV 320
Db 78 YVDPPLTV 84

RESULT 15

ABB55442 standard; Protein; 101 AA.

ABB55442;

16-MAY-2002 (first entry)

Lactococcus lactis protein rplX.

Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

Lactococcus lactis IL1403.

FR2807446-A1.

12-OCT-2001.

11-APR-2000; 2000FR-0004630.

11-APR-2000; 2000FR-0004630.

(INRG) INRA INST NAT RECH AGRONOMIQUE.

Bolotline A, Sorokine A, Renault P, Ehrlich SD;

WPI; 2002-043418/06.

New nucleotide sequence useful in the identification of Lactococcus

lactis and related species -

Claim 6; SEQ ID No 2144; 2504bp; French.

The present invention is related to a Lactococcus lactis nucleotide sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic acid sequence is useful in the detection and/or amplification of related species. The proteins of the invention are useful for the biosynthesis or biodegradation of a composition of interest. The invention helps research in lactic bacteria, particularly useful in the production of yogurt and cheese.
Note: The sequence data for this patent is based on equivalent patent WO2001/7334 (published 18-OCT-2001) which is available in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

Sequence 101 AA;

Query Match 2.2%; Score 7; DB 23; Length 101;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 279 VKTGDTV 285
Db 3 VKTGDTV 9

RESULT 16

AA658104 standard; Protein; 108 AA.

AA658104;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 74962.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0132048.

04-MAY-1999; 99US-0132407.

05-MAY-1999; 99US-0132484.

06-MAY-1999; 99US-0132485.

07-MAY-1999; 99US-0132486.

11-MAY-1999; 99US-0132487.

14-MAY-1999; 99US-0132488.

14-MAY-1999; 99US-0132489.

18-MAY-1999; 99US-0134370.

19-MAY-1999; 99US-0134768.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136352.

28-MAY-1999; 99US-0137222.

01-JUN-1999; 99US-0137528.

03-JUN-1999; 99US-0137502.

04-JUN-1999; 99US-0137724.

07-JUN-1999; 99US-0138094.

08-JUN-1999; 99US-0138540.

10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156468.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159884.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 2.2%; Score 7; DB 21; Length 108;

Matches 7; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 RERSISS 129
DB 99 RERSISS 105

AC ABG04226;
 XX 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #A217.
 XX
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS68413.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 34585; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 126 AA;
 XX
 QY Query Match 2.2%; Score 7; DB 22; Length 126;
 XX Best Local Similarity 100.0%; Pred. No. 1e+02;
 Db Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 181 VAPPVTE 187
 |||||
 112 VAPPVTE 118

DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 18970.
 XX
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukemia;
 KM nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI85009.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20; SEQ ID NO 18970; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, hematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 136 AA;
 XX
 QY Query Match 2.2%; Score 7; DB 22; Length 136;
 XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Db Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 167 KPTPTTP 173
 |||||
 28 KPTPTTP 34

RESULT 21
 AAG58103
 ID AAG58103 standard; Protein; 138 AA.
 AC AAG58103;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 74961.
 XX
 KM Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.
 XX

PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160960.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.2%; Score 7; DB 21; Length 138;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 RERSISS 129
 |||||
 DB 129 RERSISS 135

RESULT 22
 ID AAO05875 standard; Protein; 147 AA.
 XX AAO05875;
 AC
 XX 06-NOV-2001 (first entry)
 DT
 XX Human polypeptide SEQ ID NO 19767.
 DE
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 XX WO200164835-A2.
 PN
 XX 07-SEP-2001.
 PD
 XX

PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX WPI; 2001-514836/56.
 DR N-PSDB; AA185806.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders.
 XX
 PS Claim 20; SEQ ID NO 19767; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 147 AA;

Query Match 2.2%; Score 7; DB 22; Length 147;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 PAPVAPP 184
 |||||
 DB 25 PAPVAPP 31

RESULT 23
 ID AAE02227 standard; Protein; 157 AA.
 XX AAE02227;
 AC
 XX 31-JUL-2001 (first entry)
 DT
 XX Staphylococcus aureus biofilm protein, biofilm A.
 DE
 XX
 XX Biofilm A; antibacterial; bacteraemia; food poisoning; osteomyelitis;
 KM oitis media; tracheitis; epiglottitis; chryoiditis; empyema; abscess;
 KM lung; splenic; retroperitoneal; endocarditis; cerebral; renal; diarrhoea;
 KM blepharitis; conjunctivitis; keratitis; endophthalmitis; dacryocystitis;
 KM cellulitis; epidiymitis; impetigo; folliculitis; wound; myositis;
 KM septic arthritis; therapy.
 XX
 OS Staphylococcus aureus 'WCUH 29'.
 XX
 XX WO200130806-A1.
 PN
 XX 03-MAY-2001.
 PD
 XX 26-OCT-2000; 2000WO-US29460.
 PF
 XX 28-OCT-1999; 99US-0161984.
 PR
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.

XX Bryant J, Burnham MKR, Lunsford RD, Throup JP;
PI WPI; 2001-308606/32.
DR N-PSDB; AAD05669.
XX
XX New biofilm protein from *Staphylococcus*, useful e.g. in screening for
PT antibacterial agents, and related nucleic acid
PS Claim 1; Page 5; 66pp; English.

The present sequence is *Staphylococcus aureus* WCUH 29 strain biofilm protein, biofilm A. The biofilm polypeptides are useful for the diagnosis and treatment of bacterial infections specifically caused by *Staphylococcus*, especially associated with formation of biofilm in a wound or on an (in-dwelling) medical device. The polypeptide is used in the treatment of diseases such as invasive and toxicogenic diseases, bacteraemia in cancer patients. It is also used in the treatment of diseases caused by the action of exotoxins, including food poisoning, scalded skin syndrome, infections of upper respiratory tract (otitis media, bacterial tracheitis, acute epiglottitis, thyroiditis), lower respiratory (empyema, lung abscess), cardiac (infective endocarditis), gastroenteric (secretory diarrhoea, splenic abscess, retroperitoneal abscess), CNS (cerebral abscess), eye (blepharitis, conjunctivitis, keratitis, endophthalmitis, preseptal and orbital cellulitis, dacryocystitis), kidney and urinary tract (epididymitis, intrarenal and perinephric abscess, toxic shock syndrome), skin (impetigo, folliculitis, cutaneous abscess, cellulitis, wound infection, bacterial myositis) bone and joint (septic arthritis, osteomyelitis).

SO Sequence 157 AA;
SQ

Query Match 2.2%; Score 7; DB 22; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LGLIFGV 24
DB 79 LGLIFGV 85

RESULT 24
AAU66705
ID AAU66705 standard; Protein; 159 AA.
XX
XX AAU66705;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Propionibacterium acnes immunogenic protein #27601.
DE
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
OS
XX
XX WO200181581-A2.
FN
XX
XX 01-NOV-2001.
PD
XX
XX 20-APR-2001; 2001WO-US12865.
PF
XX
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71.
DR N-PSDB; AAS59749.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris
PS Example 1; SEQ ID NO 27900; 1069pp; English.

Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis. *P. acnes* is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of *P. acnes* in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for *P. acnes* proteins. These antibodies can be used to downregulate expression and activity of *P. acnes* polypeptides and therefore treat *P. acnes* infections. The antibodies may also be used as diagnostic agents for determining *P. acnes* presence, for example, by enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 159 AA;
SQ

Query Match 2.2%; Score 7; DB 22; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 VVVVKKP 168
DB 151 VVVVKKP 157

RESULT 25
AAU43362
ID AAU43362 standard; Protein; 159 AA.
XX
XX AAU43362;
AC
XX
XX 27-FEB-2002 (first entry)
DT
XX
XX Propionibacterium acnes immunogenic protein #4258.
DE
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
OS
XX
XX WO200181581-A2.
FN
XX
XX 01-NOV-2001.
PD
XX
XX 20-APR-2001; 2001WO-US12865.
PF
XX
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

```

DR WPI: 2001-616774/71.
XX N-PSDB; AAS59520.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PR treating acne vulgaris -
XX
XX PS Example 1; SEQ ID NO 4557; 1063pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 159 AA;
XX
XX Query Match 2.2%; Score 7; DB 22; Length 159;
XX Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Oy 162 VVVVVKKP 168
XX |||||
XX |||||
XX Db 151 VVVVVKKP 157
XX
XX RESULT 26
XX ABRP25478
XX ID ABRP25478 standard; Protein; 179 AA.
XX
XX AC ABRP25478;
XX
XX DE 02-JUL-2002 (first entry)
XX
XX DT Streptococcus polypeptide SEQ ID NO 132.
XX
XX DE Streptococcus polypeptide SEQ ID NO 132.
XX
XX KW Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; Infection; vaccine; meningitis; gene therapy.
XX
XX OS Streptococcus pyogenes.
XX
XX PN WO200234771-A2.
XX
XX PD 02-MAY-2002.
XX
XX PF 29-OCT-2001; 2001WO-GB04789.
XX
XX PR 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX 07-MAR-2001; 2001GB-0005640.
XX
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C,
XX Tectelin H,
XX

```

DR	WPI; 2002-352536/38.
DR	N-PSDB; ABN66109.
XX	
XX	
PT	New Streptococcus protein for the treatment or prevention of infection
PT	or disease caused by Streptococcus bacteria, such as meningitis, and
PT	for detecting a compound that binds to the protein -
XX	
XX	
PS	Claim 1; Page 3167; 4525pp; English.
XX	
CC	The invention relates to a protein (ABP25413-ABP30895) from group B
CC	streptococcus (GBS (Streptococcus agalactiae) or group A streptococcus/CAS
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC	the specification. The proteins have antibacterial and antiinflammatory
CC	activity. (I), nucleic acids encoding (I), ABN6604-ABN71526 and
CC	antibodies that bind (I) are used in the manufacture of medicaments for
CC	the treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a
CC	biological sample. (I) is used to determine whether a compound binds to
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (I) may be used to recombinantly produce (I) and may be
CC	used in gene therapy. Antibodies to (I) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
CC	Streptococcus proteins.
XX	
SQ	Sequence 179 AA;
Query Match	2.2%; Score 7; DB 23; Length 179;
Best Local Similarity	100.0%; Pred. No. 1.4e+02;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	52 GSGGGLAI 58
	51 GSGGGLAI 57
Db	
RESULT 27	
ID	AA75693
AA75693	standard; Protein, 185 AA.
XX	
XX	AA75693;
XX	
DT	21-MAR-2000 (first entry)
XX	
DE	Neisseria gonorrhoeae ORF 931 protein sequence SEQ ID NO:2858.
XX	
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW	antibacterial; gene therapy.
XX	
OS	Neisseria gonorrhoeae.
XX	
FN	W09957280-A2.
XX	
PD	11-NOV-1999.
XX	
PF	30-APR-1999; 99WO-US09346.
XX	
PR	01-MAY-1998; 98US-0083758.
PR	31-JUL-1998; 98US-0094869.
PR	02-SEP-1998; 98US-0089994.
PR	02-SEP-1998; 98US-0099062.
PR	09-OCT-1998; 98US-0103749.
PR	09-OCT-1998; 98US-0103794.
PR	09-OCT-1998; 98US-0103796.
PR	25-FEB-1999; 99US-0121528.
XX	
PA	(CHIR) CHIRON CORP.
PA	(GENO-) INST GENOMIC RES.
XX	
XX	Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M,

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarsella M;
 PI Tettelin H, Venter JC;
 XX WPI: 2000-062150/05.
 DR N-PSDB; AAZ54455.
 XX
 PT Novel Neisseria polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 1343; 1453pp; English.
 XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 SQ Sequence 185 AA;
 Query Match 2.2%; Score 7; DB 21; Length 185;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 79 DTVSKIA 85
 Db 152 DTVSKIA 158
 RESULT 28
 ABB61524
 ID ABB61524 standard; Protein; 195 AA.
 AC ABB61524;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 11364.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US092331.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 DR N-PSDB; AB05627.
 DR WPI: 2001-656660/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 11364; 21pp + Sequence Listing; English.
 XX

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AAB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 195 AA;
 Query Match 2.2%; Score 7; DB 22; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 51 SGGGGLA 57
 Db 179 SGGGGLA 185
 RESULT 29
 AAW71722
 ID AAW71722 standard; Protein; 208 AA.
 AC AAW71722;
 XX
 DT 11-JAN-1999 (first entry)
 DE Arabidopsis leafy-cotyledon 1 (LEC1) polypeptide.
 KW LEC1; leafy-cotyledon 1; embryo; transcription factor;
 KW transgenic plant.
 XX
 OS Arabidopsis thaliana ecotype Wassilewskija.
 XX
 FH Key Location/Qualifiers
 FT Binding-site 34..40
 FT FT /note= "putative DNA binding site"
 FT Domain 61..72
 FT /note= "putative subunit interaction domain"
 XX
 PN WO9637184-A1.
 XX
 PD 27-AUG-1998.
 XX
 PF 20-FEB-1998; 98WO-US02998.
 XX
 PR 19-FEB-1998; 98US-0804534.
 PR 21-FEB-1997; 97US-0804534.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Fischer RL, Goldberg RB, Harada JT, Lotan T, Ohto M;
 XX
 DR WPI: 1998-531499/45.
 DR N-PSDB; AAV61022-24.
 XX
 PT New isolated plant leafy cotyledon-1 gene - used to develop products
 PT for, e.g. increasing storage protein content in plant tissues, or
 PT producing fruit with small and non-viable seed
 XX
 PS Claim 15; Page 31; 55pp; English.
 XX
 CC LEC1 polypeptide is encoded by the leafy cotyledon-1 (LEC1) gene
 CC (see AAV61022-24) of Arabidopsis thaliana. Full-length LEC1
 CC polypeptide can act as a subunit of a protein capable of acting
 CC as a transcription factor in plant cells. The LEC1 gene is
 CC embryo-specific and can be used to modulate development (claimed) of
 CC embryos or other organs in plants. Inhibiting expression can be
 CC useful e.g. in weed control (by transferring an inhibitory sequence

CC to a weedy species and allowing it to be transmitted through sexual
CC crosses) or to produce fruit with small and non-viable seed.
CC Enhanced expression of LEC1 can be used to increase storage protein
CC content in plant tissues to improve nutritional value.

SO Sequence 208 AA;

Query Match 2.2%; Score 7; DB 19; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 YVDP LTV 320
|||
105 YVDP LTV 111

RESULT 30
AAV54563 standard; Protein; 208 AA.

AC AAV54563;

DT 25-APR-2000 (first entry)

XX Amino acid sequence of an Arabidopsis leafy cotyledon 1 polypeptide.

XX Leafy cotyledon 1 gene; LEC1 gene; embryogenesis; cotyledon identity;
KW embryo development; transcription factor; seed development;
KW ecotopic development; embryonic plant tissue; weed control;
KW nutritional value; storage protein; cotyledon; seed;
KW reproductive tissue mass; fruit size; seed mass.

OS Arabidopsis thaliana.

PN W09967405-A2.

PD 29-DEC-1999.

PF 24-JUN-1999; 99WO-US14384.

PR 24-JUN-1998; 98US-0103478.

PR 17-NOV-1998; 98US-0193931.

PA (REGC) UNIV CALIFORNIA.

PI Harada J, Lotan T, Ohto M, Goldberg RB, Fischer RL;

DR WPI; 2000-160588/14.

DR N-PSDB; AAZ45838.

PT New embryo-specific gene useful for producing transgenic plant
XX Claim 6; Page 53-54; 69pp; English.

XX The present sequence represents a leafy cotyledon 1 (LEC1) polypeptide.
CC LEC1 genes are thought to play a central role in late embryogenesis, in
CC specifying cotyledon identity during embryo development. LEC1
CC polypeptides may act as transcription factors. LEC1 polynucleotides are
CC useful for modulating seed development and for inducing ecotopic
CC development of embryonic tissue in a plant. In both cases, the LEC1
CC polynucleotide is introduced into the plant through a sexual cross and
CC is co-expressed in an antisense orientation with a second heterologous
CC polynucleotide selected from AP2 and RAR2 genes of Arabidopsis. The
CC LEC1 polynucleotide is also useful for targeting expression in a seed,
CC and for preparing expression cassettes for suppressing or enhancing
CC endogenous LEC1 gene expression, which is useful in weed control or
CC for improving nutritional value of plant tissue respectively. LEC1
CC polypeptides and polynucleotides are especially used for increasing or
CC decreasing storage protein content in cotyledons or leaves. LEC1 also
CC increases reproductive tissue mass, e.g., increases fruit size, seed
CC mass, its protein or its oil.

XX Sequence 208 AA;

Query Match 2.2%; Score 7; DB 21; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 YVDP LTV 320
|||
105 YVDP LTV 111

RESULT 31
AAE09828 standard; Protein; 208 AA.

AC AAE09828;

DT 29-NOV-2001 (first entry)

XX Leafy cotyledon 1 (LEC1) protein.

XX Leafy cotyledon 1; LEC1; transcription modulator; seed development;
KW embryonic characteristic; transgenic plant; storage protein content;
KW weed control; ecotopic expression.

XX Unidentified.

XX Key Location/Qualifiers
FT Domain 28..117
FT /label=B-domain

PN W0200164022-A2.

PD 07-SEP-2001.

PF 21-FEB-2001; 2001WO-US05454.

PR 01-MAR-2000; 2000US-0516052.

PA (REGC) UNIV CALIFORNIA.

PI Harada J, Lotan T, Ohto M, Goldberg RB, Fischer RL, Bui A;

PI Kwong R;

DR WPI; 2001-565462/63.

DR N-PSDB; AAD16914, AAD16915, AAD16916.

PT Novel leafy cotyledon 1 polynucleotide encoding leafy cotyledon
PT polypeptide for modulating transcription resulting in seed development
PT in plant comprises providing in an expression cassette linked to LEC1
PT gene promoter -

PS Claim 2; Page 53-54; 73pp; English.

XX The patent discloses novel leafy cotyledon 1 (LEC1) polypeptides and
CC polynucleotides encoding them. The invention also provides expression
CC cassettes comprising a promoter operably linked to a heterologous
CC sequence that encodes LEC-1 protein comprising a subsequence that
CC is similar to LEC-1 B domain. LEC-1 sequences are used for modulating
CC transcription resulting in the induction of embryonic characteristics
CC or seed development in a plant. Polynucleotide sequences from LEC1
CC genes are used to direct expression of desired heterologous genes in
CC embryos, to modulate the development of embryos or the characteristics
CC of an embryo on other organs (e.g., by enhancing expression of the gene
CC in a transgenic plant). Modulation of the expression of LEC-1 gene is
CC used to manipulate a number of useful traits, such as increasing or
CC decreasing storage protein content in cotyledons or leaves. Inhibition
CC of LEC1 gene expression is useful in weed control or to produce fruit
CC with small and non-viable seed. Enhanced expression of LEC1 genes is
CC useful to increase storage protein content in plant tissues. The LEC-1
CC gene promoters are useful for directing gene expression so that the
CC desired gene product is located in embryos or seeds. Expression of
CC LEC-1 genes in plant reproductive or vegetative tissue induce ectopic
CC expression of cells and tissues, together with another plant nucleic

[illegible]

PR	17-NOV-2000;	2000US-02492217.	
PR	17-NOV-2000;	2000US-02492218.	
PR	17-NOV-2000;	2000US-02492244.	
PR	17-NOV-2000;	2000US-02492245.	
PR	17-NOV-2000;	2000US-02492264.	
PR	17-NOV-2000;	2000US-02492265.	
PR	17-NOV-2000;	2000US-02492297.	
PR	17-NOV-2000;	2000US-02492299.	
PR	17-NOV-2000;	2000US-02493000.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
PA	(HMDA-) HUMAN GENOME SCI INC.		
XX			
PI	Rosen CA, Barash SC, Ruben SM;		
DR	WP1; 2001-465572/50.		
DR	N-PSDB; AAS31221.		
XX			
PT	Nucleic acid molecules encoding human secreted extracellular matrix proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -		
XX			
PS	Claim 11; SEQ ID No 300; 577pp; English.		
XX			
CC	The invention relates to isolated nucleic acid molecules encoding novel human secreted extracellular matrix proteins (SPs). The polynucleotides and proteins are used to prevent, treat a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For example, disorders associated with decreased expression of SPs. The SP polynucleotide or a vector expressing them may be administered to treat diseases by gene therapy. Antisense molecules may be administered to down regulate expression of SPs by binding with the cells own genes and preventing their expression. The polynucleotides may also be used as DNA probes in diagnostic assays. The SPs may also be used as antigens to produce antibodies and to identify modulators (agonists and antagonists) of the SPs. The anti-(SP) antibodies and antagonists may also be used to down regulate expression and activity of SP and as diagnostic agents for detecting the presence of SPs in samples. The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). Other uses include wound healing, maintenance of organs before transplantation,		
CC			
Query Match	2.2%	Score 7;	DB 22; Length 251;
Best Local Similarity	100.0%;	Pred. No. 2e+02;	
Matches	7;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Cy	167	KPPTPP	173
Db	211	KPPTPP	217
RESULT 34			
ABP47870			
ID	ABP47870 standard; Protein; 251 AA.		
XX			

AC ABP47870;
XX 23-AUG-2002 (first entry)
DT
XX
DE Human polypeptide SEQ ID NO 300.
XX
XX Human; noctropic; neuroprotective; cytosolic; dermatological; virucide;
XX immunosuppressive; anti-inflammation; anti-HIV; antibacterial; vulnery;
XX antiparkinsonian; antischistosomal; antianemic; antitubercular; cancer;
XX antidiabetic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antituber; anticonvulsant; antifungal;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; neurotropic; gene therapy; vaccine.
OS
XX Homo sapiens.
XX
XX US2002042386-A1.
XX
XX 11-APR-2002.
XX
XX 17-JAN-2001; 2001US-0764870.
XX
XX 31-JAN-2000; 2000US-179065P.
XX 04-FEB-2000; 2000US-180628P.
XX 28-JUN-2000; 2000US-214868P.
XX 07-JUL-2000; 2000US-216647P.
XX 07-JUL-2000; 2000US-216680P.
XX 11-JUL-2000; 2000US-217487P.
XX 11-JUL-2000; 2000US-218290P.
XX 26-JUL-2000; 2000US-220963P.
XX 26-JUL-2000; 2000US-220964P.
XX 14-AUG-2000; 2000US-224518P.
XX 14-AUG-2000; 2000US-224519P.
XX 14-AUG-2000; 2000US-225267P.
XX 14-AUG-2000; 2000US-225268P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225477P.
XX 14-AUG-2000; 2000US-225757P.
XX 14-AUG-2000; 2000US-225758P.
XX 22-AUG-2000; 2000US-226868P.
XX 30-AUG-2000; 2000US-228924P.
XX 01-SEP-2000; 2000US-229287P.
XX 01-SEP-2000; 2000US-229343P.
XX 01-SEP-2000; 2000US-229344P.
XX 05-SEP-2000; 2000US-229509P.
XX 05-SEP-2000; 2000US-229513P.
XX 08-SEP-2000; 2000US-231413P.
XX 21-SEP-2000; 2000US-234233P.
XX 21-SEP-2000; 2000US-234274P.
XX 25-SEP-2000; 2000US-234997P.
XX 27-SEP-2000; 2000US-235834P.
XX 29-SEP-2000; 2000US-236377P.
XX 29-SEP-2000; 2000US-236378P.
XX 29-SEP-2000; 2000US-236388P.
XX 29-SEP-2000; 2000US-236399P.
XX 29-SEP-2000; 2000US-236370P.
XX 02-OCT-2000; 2000US-236802P.
XX 02-OCT-2000; 2000US-237037P.
XX 02-OCT-2000; 2000US-237038P.
XX 02-OCT-2000; 2000US-237039P.
XX 02-OCT-2000; 2000US-237040P.
XX 13-OCT-2000; 2000US-239935P.
XX 20-OCT-2000; 2000US-240960P.
XX 20-OCT-2000; 2000US-241785P.
XX 20-OCT-2000; 2000US-241809P.
XX 01-NOV-2000; 2000US-244617P.
XX 17-NOV-2000; 2000US-249299P.
XX 08-DEC-2000; 2000US-251856P.
XX 08-DEC-2000; 2000US-251868P.
XX 08-DEC-2000; 2000US-251869P.

PA (ROSE/) ROSEN C A.
PA (RUBEN/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX WPI, 2002-470713/50.
XX N-PsDB; ABQ66545.
XX
XX New nucleic acid encoding human proteins, useful for diagnosis,
XX treatment and prevention of e.g. osteoporosis, also related
XX polypeptides and antibodies -
XX
XX Claim 11; SEQ ID NO 300; 235pp + Sequence Listing; English.
XX
XX
XX The invention relates to novel genes (ABQ66521-ABQ66785) and proteins
XX (ABP47846-ABP48110) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (anti)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia and
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from USPTO at seqdata.uspto.gov/sequence.html?docID=99909764870.
XX
XX
XX SQ Sequence 251 AA;
XX
XX
XX Query Match 2.2%; Score 7; DB 23; Length 251;
XX Best Local Similarity 100.0%; Pred. No. 2e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 167 KPTPTP 173
XX |||||
XX 211 KPTPTP 217
XX
XX
XX RESULT 35
XX ID ABP30505 standard; Protein; 298 AA.
XX
XX ABP30505;
XX
XX 02-JUL-2002 (first entry)
XX
XX Streptococcus polypeptide SEQ ID NO 10186.
XX
XX Streptococcus agalactiae.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX

XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tetteijn H;
 XX WPI; 2002-352536/38.
 DR N-PSDB; ABN71136.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 or disease caused by Streptococcus bacteria, such as meningitis, and
 for detecting a compound that binds to the protein -
 XX
 PS Claim 1; Page 4148; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins..
 XX
 SQ Sequence 298 AA;
 XX
 Query Match 2.2%; Score 7; DB 23; Length 298;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 63 ITDSQGV 69
 DB 24 ITDSQGV 30
 XX
 RESULT 36
 ABB08497
 ID ABB08497 standard; Protein; 300 AA.
 XX
 AC ABB08497;
 XX
 DT 23-MAY-2002 (first entry)
 XX
 DE Envelope 2 protein of HCV.
 XX
 KW Envelope 2 protein; HCV.
 XX
 OS Unidentified.
 XX
 PN KR138597-B1.
 XX
 PD 30-APR-1998.
 XX
 PF 30-JUL-1994; 94KR-0018832.
 XX
 PR 30-JUL-1994; 94KR-0018832.
 XX
 PA (GLDS) LG CHEM CO LTD.
 XX
 PI Ryu W, Yang J, Cho J;
 XX
 DR WPI; 2000-144176/13.
 DR N-PSDB; ABL41548.
 XX
 PT STABILISED RECOMBINANT ANIMAL CELL LINE EXPRESSING ENVELOPE 2 PROTEIN
 OF HCV

XX Disclosure; Page 7; 8pp; Korean.
 XX
 PS This invention relates to stabilised recombinant animal cell line
 CC expressing envelope 2 protein of HCV. The present sequence
 CC represents the envelope 2 protein of HCV.
 XX
 SQ Sequence 300 AA;
 XX
 Query Match 2.2%; Score 7; DB 21; Length 300;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 136 TPSPVAV 142
 DB 141 TPSPVAV 147
 XX
 RESULT 37
 AAR98460
 ID AAR98460 standard; Protein; 304 AA.
 XX
 AC AAR98460;
 XX
 DT 18-DEC-1996 (first entry)
 XX
 DE Endonuclease LlaIIc.
 XX
 KW LlaIIA; LlaIIB; LlaIIC; m6a-methyltransferase; endonuclease; restriction;
 KW modification system; phage resistance; milk; dairy product;
 KW food additive.
 XX
 OS Lactococcus lactis SMQ-17.
 XX
 PN W09621017-A2.
 XX
 PD 11-JUL-1996..
 XX
 PF 29-DEC-1995; 95WO-NL00448.
 XX
 PR 19-APR-1995; 95US-0424641.
 PR 30-DEC-1994; 94US-0366480.
 XX
 PA (UNIL) QUEST INT BV.
 XX
 PI Moineau S, Vandenberg PA, Vedamuthu ER, Walker SA;
 XX
 DR WPI; 1996-333992/33.
 DR N-PSDB; AAT30400.
 XX
 PT DNA encoding restriction and modification enzymes from L. lactis -
 PT confers phage resistance to bacteria, for use in milk and as a food
 PT additive
 XX
 PS Claim 2; Fig 4; 48pp; English.
 XX
 CC The present sequence is that of the LlaIIC gene product, an
 CC endonuclease, part of a Lactococcus lactis restriction/modification (R/M)
 CC system. The LlaII R/M system is similar to the MboI system which
 CC recognises the sequence 5'-GATC-3' and cleaves it before the guanine. A
 CC plasmid contg. the DNA (AAT30400) encoding the R/M system can confer
 CC phage resistance to a bacterium, such as providing resistance to
 CC phage-sensitive dairy cultures. The DNA and proteins can be used for
 CC improving fermentation of a dairy prod. and as additives in food
 CC processing.
 CC
 SQ Sequence 304 AA;
 XX
 Query Match 2.2%; Score 7; DB 17; Length 304;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 80 TVSKIAQ 86

Db 171 TVSKIAQ 177

RESULT 38

XX ID AAW02163 standard; Protein; 304 AA.

XX AC AAW02163;

XX DT 08-DEC-1996 (first entry)

XX DE Lactococcus lactis ssp. cremoris W9 restriction endonuclease R.LIAAI.

XX KM Lactic acid bacterium; Danish starter culture; cheese; enzyme;

XX OS restriction endonuclease; methylase; fermented milk; phage resistance.

XX OS Lactococcus lactis subsp. cremoris W9.

XX MO9625503-A1.

XX 22-AUG-1996.

XX PF 19-FEB-1996; 96WO-DK00076.

XX PR 17-FEB-1995; 95DK-0000179.

XX PA (JOSEF) JOSEPHSEN J.

XX PA (MADS/) MADSEN A.

XX PA (NYEN/) NYENGAARD N.R.

XX PA (VOGE/) VOGENSEN F.K.

XX PI Josephsen J, Madsen A, Nyengaard NR, Vogensen FK;

XX DR WPI; 1996-393404/39.

XX DR N-PSDB; AAT36391.

XX PT Type II restriction-modification systems obtd. from Lactococcus

XX PT lactis - for conferring phage resistance on lactic acid bacteria,

XX PT useful as starter cultures for cheese and fermented milk prods.

XX PS Claim 1; Page 65-66; 93pp; English.

XX CC This AA sequence encodes a restriction endonuclease, R.LIAAI, with

XX CC the recognition sequence 5'-GATC-3' (* indicating point of

XX CC cleavage) from the plasmid-derived type-II restriction-modification

XX CC system from Lactococcus lactis subsp. cremoris W9, LIAAI. R.LIAAI is

XX CC one of 3 proteins in this restriction-modification system, along with

XX CC a 284 AA methylase (M.LIAAI; AAW02160), and a 269 or 257 AA methylase

XX CC (M.LIAAI; AAW02161 and AAW02162, respectively). This restriction-

XX CC modification system may be used in a method for conferring increased

XX CC virus resistance, more specifically phage resistance, to a L. lactis

XX CC strain used in cheese manufacture.

XX SQ Sequence 304 AA;

XX RESULT 39

XX ID ABP28875 standard; Protein; 321 AA.

XX AC ABP28875;

XX DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 6926.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX OS Streptococcus agalactiae.

XX MO200234771-A2.

XX 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;

XX PI Tettelin H;

XX DR WPI; 2002-352536/38.

XX DR N-PSDB; AEN69506.

XX PT New streptococcus protein for the treatment or prevention of infection

XX PT or disease caused by Streptococcus bacteria, such as meningitis, and

XX PT for detecting a compound that binds to the protein -

XX PS Claim 1; Page 3855; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B

XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

XX CC the specification. The proteins have antibacterial and antiinflammatory

XX CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

XX CC antibodies that bind (I) are used in the manufacture of medicaments for

XX CC the treatment or prevention of infection or disease caused by

XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a

XX CC biological sample. (I) is used to determine whether a compound binds to

XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

XX CC used as a vaccine or diagnostic composition. The disease caused by

XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic

XX CC acid encoding (I) may be used to recombinantly produce (I) and may be

XX CC used in gene therapy. Antibodies to (I) are used for affinity

XX CC chromatography, immunoassays, and distinguishing/identifying

XX CC Streptococcus proteins.

XX SQ Sequence 321 AA;

XX Query Match 2.2%; Score 7; DB 23; Length 321;

XX Best Local Similarity 100.0%; Pred. No. 2.5e+02; Mismatches 0; Gaps 0;

XX Matches 7; Conservative 0; Indels 0; Gaps 0;

XX Db 63 ITDSQGV 69

XX 47 ITDSQGV 53

XX RESULT 40

XX ID AAG38356 standard; Protein; 337 AA.

XX AC AAG38356;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 47308.

XX Protein identification; signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EPI033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-011825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 16-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.

24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

Wed Jul 9 10:00:57 2003

us-10-018-706-2.01go.rag

Page 32

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161407.
PR 25-OCT-1999; 99US-0161408.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.2%; Score 7; DB 21; Length 337;
Best Local Similarity 100.0%; Pred.No.2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 VRSIS 128
Db 129 VRSIS 135

Search completed: July 8, 2003, 11:18:50
Job time : 95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 8, 2003, 11:17:18 / Search time 27 Seconds
(without alignments)
350.896 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 322
Sequence: 1 MVTAINSONQKPIKRLGL.....LFEFRISRMGVYDPLTLK 322

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Issued Patente_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.5	557	4	US-09-027-064-2
2	8	2.5	557	4	US-09-271-815-2
3	7	2.2	13	1	US-08-258-152-9
4	7	2.2	13	2	US-08-076-299A-9
5	7	2.2	13	2	US-08-438-582-9
6	7	2.2	13	4	US-09-256-596-9
7	7	2.2	13	4	US-08-479-737-44
8	7	2.2	13	4	US-08-475-442A-44
9	7	2.2	56	3	US-09-029-424-15
10	7	2.2	90	4	US-09-103-478-19
11	7	2.2	90	4	US-09-193-931C-19
12	7	2.2	208	4	US-09-103-478-2
13	7	2.2	208	4	US-09-193-931C-2
14	7	2.2	304	2	US-08-424-641B-4
15	7	2.2	304	2	US-08-820-980-4
16	7	2.2	304	2	US-08-826-439-4
17	7	2.2	304	4	US-08-913-159-8
18	7	2.2	344	4	US-09-147-236-11
19	7	2.2	350	3	US-08-909-828-1
20	7	2.2	350	3	US-08-909-828-2
21	7	2.2	356	4	US-09-232-191-4
22	7	2.2	356	4	US-09-232-200-4
23	7	2.2	356	4	US-09-232-197-4
24	7	2.2	356	4	US-09-232-201-4
25	7	2.2	501	1	US-08-149-105-15
26	7	2.2	501	1	US-08-317-847-15
27	7	2.2	505	4	US-09-382-256-8

28	7	2.2	505	4	US-09-382-256-16	Sequence 16, Appl
29	7	2.2	505	4	US-09-395-115-8	Sequence 8, Appl
30	7	2.2	505	4	US-09-395-115-16	Sequence 16, Appl
31	7	2.2	505	4	US-08-123-934A-8	Sequence 8, Appl
32	7	2.2	505	4	US-08-436-265-8	Sequence 8, Appl
33	7	2.2	505	4	US-08-436-265-16	Sequence 16, Appl
34	7	2.2	505	4	US-09-679-187-8	Sequence 8, Appl
35	7	2.2	505	4	US-09-679-187-16	Sequence 16, Appl
36	7	2.2	505	5	PCT-US94-10080-8	Sequence 8, Appl
37	7	2.2	543	2	US-08-469-412A-7	Sequence 7, Appl
38	7	2.2	543	2	US-09-021-715-7	Sequence 7, Appl
39	7	2.2	623	4	US-09-232-200-9	Sequence 89, Appl
40	7	2.2	623	4	US-09-232-200-98	Sequence 89, Appl
41	7	2.2	623	4	US-09-232-197-89	Sequence 89, Appl
42	7	2.2	623	4	US-09-232-197-88	Sequence 89, Appl
43	7	2.2	623	4	US-09-232-201-89	Sequence 89, Appl
44	7	2.2	623	4	US-09-232-201-98	Sequence 98, Appl
45	7	2.2	843	4	US-09-235-451-25	Sequence 25, Appl
46	7	1.9	8	2	US-08-803-899-18	Sequence 18, Appl
47	6	1.9	8	3	US-09-029-424-19	Sequence 19, Appl
48	6	1.9	9	1	US-08-146-145-3	Sequence 3, Appl
49	6	1.9	9	3	US-08-481-985B-130	Sequence 130, App
50	6	1.9	10	3	US-08-476-509B-31	Sequence 31, Appl
51	6	1.9	11	2	US-08-803-899-20	Sequence 20, Appl
52	6	1.9	12	2	US-08-803-899-19	Sequence 19, Appl
53	6	1.9	13	2	US-08-803-899-12	Sequence 12, Appl
54	6	1.9	13	2	US-08-803-899-17	Sequence 17, Appl
55	6	1.9	14	1	US-08-061-092A-11	Sequence 11, Appl
56	6	1.9	14	1	US-08-481-003-31	Sequence 31, Appl
57	6	1.9	14	2	US-08-256-156A-1	Sequence 1, Appl
58	6	1.9	14	2	US-08-373-190-6	Sequence 6, Appl
59	6	1.9	14	2	US-08-224-591-8	Sequence 8, Appl
60	6	1.9	14	2	US-08-392-138A-2	Sequence 2, Appl
61	6	1.9	14	2	US-08-438-190A-6	Sequence 6, Appl
62	6	1.9	14	2	US-08-926-789-8	Sequence 8, Appl
63	6	1.9	14	2	US-08-818-253-12	Sequence 42, Appl
64	6	1.9	14	2	US-08-818-253-14	Sequence 44, Appl
65	6	1.9	14	3	US-08-350-215-6	Sequence 6, Appl
66	6	1.9	14	3	US-09-166-750-2	Sequence 2, Appl
67	6	1.9	14	3	US-09-166-093-2	Sequence 2, Appl
68	6	1.9	14	3	US-09-287-145A-6	Sequence 6, Appl
69	6	1.9	14	3	US-08-485-598-31	Sequence 31, Appl
70	6	1.9	14	3	US-08-454-098-23	Sequence 23, Appl
71	6	1.9	14	3	US-09-172-019-2	Sequence 2, Appl
72	6	1.9	14	3	US-09-166-094-2	Sequence 2, Appl
73	6	1.9	14	4	US-08-501-253A-10	Sequence 10, Appl
74	6	1.9	14	4	US-08-818-252-42	Sequence 42, Appl
75	6	1.9	14	4	US-08-818-252-44	Sequence 44, Appl
76	6	1.9	14	4	US-09-362-805-4	Sequence 4, Appl
77	6	1.9	14	4	US-09-362-805-6	Sequence 6, Appl
78	6	1.9	14	4	US-09-173-190-4	Sequence 4, Appl
79	6	1.9	14	4	US-09-173-190-6	Sequence 6, Appl
80	6	1.9	14	4	US-09-069-821-7	Sequence 7, Appl
81	6	1.9	14	4	US-09-556-111-6	Sequence 6, Appl
82	6	1.9	14	4	US-09-316-919-58	Sequence 58, Appl
83	6	1.9	14	4	US-09-316-919-60	Sequence 60, Appl
84	6	1.9	14	5	PCT-US93-11138-8	Sequence 8, Appl
85	6	1.9	14	5	PCT-US96-01600-23	Sequence 23, Appl
86	6	1.9	15	2	US-08-687-956A-3	Sequence 3, Appl
87	6	1.9	15	2	US-08-687-956A-6	Sequence 6, Appl
88	6	1.9	18	2	US-08-224-591-1	Sequence 1, Appl
89	6	1.9	18	2	US-08-224-591-9	Sequence 9, Appl
90	6	1.9	18	2	US-08-224-591-10	Sequence 10, Appl
91	6	1.9	18	2	US-08-332-338A-3	Sequence 3, Appl
92	6	1.9	18	2	US-08-819-033-3	Sequence 3, Appl
93	6	1.9	18	2	US-08-926-789-1	Sequence 1, Appl
94	6	1.9	18	2	US-08-926-789-9	Sequence 9, Appl
95	6	1.9	18	2	US-08-926-789-10	Sequence 10, Appl
96	6	1.9	18	2	US-08-818-253-43	Sequence 43, Appl
97	6	1.9	18	2	US-08-818-253-45	Sequence 45, Appl
98	6	1.9	18	3	US-09-166-750-3	Sequence 3, Appl
99	6	1.9	18	3	US-09-166-093-3	Sequence 3, Appl
100	6	1.9	18	3	US-09-172-019-3	Sequence 3, Appl

539	5	1.6	11	4	US-09-001-984C-23	Sequence 23, Appl	612	5	1.6	16	4	US-09-500-124-392	Sequence 392, App
540	5	1.6	11	4	US-09-001-984C-55	Sequence 55, Appl	613	5	1.6	16	4	US-09-170-769A-11	Sequence 11, Appl
541	5	1.6	12	1	US-08-423-399B-13	Sequence 13, Appl	614	5	1.6	17	1	US-07-920-519-21	Sequence 21, Appl
542	5	1.6	12	1	US-08-423-399B-24	Sequence 24, Appl	615	5	1.6	17	1	US-08-086-410-18	Sequence 18, Appl
543	5	1.6	12	1	US-08-446-856A-10	Sequence 10, Appl	616	5	1.6	17	1	US-08-314-586-21	Sequence 21, Appl
544	5	1.6	12	1	US-08-446-856A-11	Sequence 11, Appl	617	5	1.6	17	1	US-08-645-930A-2	Sequence 2, Appl
545	5	1.6	12	2	US-08-224-591-7	Sequence 7, Appl	618	5	1.6	17	2	US-08-484-905-10	Sequence 10, Appl
546	5	1.6	12	2	US-08-392-338A-1	Sequence 1, Appl	619	5	1.6	17	3	US-08-481-985B-10	Sequence 10, Appl
547	5	1.6	12	2	US-08-926-789-7	Sequence 7, Appl	620	5	1.6	17	3	US-08-785-247-10	Sequence 10, Appl
548	5	1.6	12	2	US-08-538-711A-2	Sequence 2, Appl	621	5	1.6	17	4	US-08-370-476-10	Sequence 10, Appl
549	5	1.6	12	2	US-08-818-253-41	Sequence 41, Appl	622	5	1.6	17	4	US-08-630-915A-1	Sequence 1, Appl
550	5	1.6	12	3	US-09-166-750-1	Sequence 1, Appl	623	5	1.6	17	4	US-09-450-972-3	Sequence 3, Appl
551	5	1.6	12	3	US-09-166-093-1	Sequence 1, Appl	624	5	1.6	17	4	US-09-101-751A-22	Sequence 22, Appl
552	5	1.6	12	3	US-09-172-019-1	Sequence 1, Appl	625	5	1.6	18	1	US-08-250-464-1	Sequence 1, Appl
553	5	1.6	12	3	US-09-166-094-1	Sequence 1, Appl	626	5	1.6	18	1	US-08-602-999A-320	Sequence 320, App
554	5	1.6	12	4	US-08-818-252-41	Sequence 41, Appl	627	5	1.6	18	4	US-09-426-680-24	Sequence 24, Appl
555	5	1.6	12	4	US-08-895-590-34	Sequence 34, Appl	628	5	1.6	18	4	US-09-500-124-320	Sequence 320, App
556	5	1.6	12	4	US-08-725-027-2	Sequence 2, Appl	629	5	1.6	19	1	US-07-976-358-10	Sequence 10, Appl
557	5	1.6	12	4	US-09-362-805-3	Sequence 3, Appl	630	5	1.6	19	1	US-07-976-358-11	Sequence 11, Appl
558	5	1.6	12	4	US-09-173-190-3	Sequence 3, Appl	631	5	1.6	19	1	US-08-244-116B-6	Sequence 6, Appl
559	5	1.6	12	4	US-09-069-821-6	Sequence 6, Appl	632	5	1.6	19	2	US-08-648-398-16	Sequence 16, Appl
560	5	1.6	12	4	US-09-044-781A-16	Sequence 16, Appl	633	5	1.6	19	3	US-08-792-832A-34	Sequence 34, Appl
561	5	1.6	12	4	US-09-513-783A-102	Sequence 102, App	634	5	1.6	19	3	US-08-908-633C-57	Sequence 57, Appl
562	5	1.6	12	4	US-09-316-919-57	Sequence 57, Appl	635	5	1.6	19	4	US-07-861-458C-104	Sequence 104, App
563	5	1.6	12	5	PCT-US93-11138-7	Sequence 7, Appl	636	5	1.6	19	4	US-09-149-476-676	Sequence 676, App
564	5	1.6	13	2	US-08-803-899-16	Sequence 16, Appl	637	5	1.6	19	4	US-09-441-502B-6	Sequence 6, Appl
565	5	1.6	13	3	US-08-782-480-93	Sequence 93, Appl	638	5	1.6	19	4	US-09-441-502B-27	Sequence 27, Appl
566	5	1.6	13	4	US-08-525-539A-66	Sequence 66, Appl	639	5	1.6	19	4	US-09-441-502B-29	Sequence 29, Appl
567	5	1.6	13	4	US-09-005-167A-93	Sequence 93, Appl	640	5	1.6	19	4	US-09-441-502B-35	Sequence 35, Appl
568	5	1.6	14	2	US-08-373-190-3	Sequence 3, Appl	641	5	1.6	19	4	US-09-441-502B-37	Sequence 37, Appl
569	5	1.6	14	2	US-08-373-190-5	Sequence 5, Appl	642	5	1.6	19	4	US-09-441-502B-43	Sequence 43, Appl
570	5	1.6	14	2	US-08-687-956A-4	Sequence 4, Appl	643	5	1.6	19	4	US-09-441-502B-48	Sequence 48, Appl
571	5	1.6	14	2	US-08-438-190A-3	Sequence 3, Appl	644	5	1.6	19	4	US-09-441-502B-49	Sequence 49, Appl
572	5	1.6	14	2	US-08-438-190A-5	Sequence 5, Appl	645	5	1.6	19	4	US-09-441-502B-55	Sequence 55, Appl
573	5	1.6	14	2	US-08-818-253-46	Sequence 46, Appl	646	5	1.6	19	4	US-09-441-502B-77	Sequence 77, Appl
574	5	1.6	14	3	US-08-350-215-3	Sequence 3, Appl	647	5	1.6	19	4	US-09-441-502B-8	Sequence 8, Appl
575	5	1.6	14	3	US-08-350-215-5	Sequence 5, Appl	648	5	1.6	19	4	US-09-441-502B-83	Sequence 83, Appl
576	5	1.6	14	3	US-09-287-145A-3	Sequence 3, Appl	649	5	1.6	19	4	US-09-441-502B-96	Sequence 96, Appl
577	5	1.6	14	3	US-09-287-145A-5	Sequence 5, Appl	650	5	1.6	20	2	US-08-934-915-19	Sequence 19, Appl
578	5	1.6	14	4	US-09-046-992-6	Sequence 6, Appl	651	5	1.6	20	2	US-08-934-915-95	Sequence 95, Appl
579	5	1.6	14	4	US-08-818-252-46	Sequence 46, Appl	652	5	1.6	20	2	US-08-934-915-95	Sequence 95, Appl
580	5	1.6	14	4	US-09-362-805-8	Sequence 8, Appl	653	5	1.6	20	3	US-08-053-451B-167	Sequence 167, App
581	5	1.6	14	4	US-09-173-190-8	Sequence 8, Appl	654	5	1.6	20	4	US-09-258-934-44	Sequence 44, Appl
582	5	1.6	14	4	US-09-193-214C-6	Sequence 6, Appl	655	5	1.6	20	4	US-08-612-973-80	Sequence 80, Appl
583	5	1.6	14	4	US-09-353-332-3	Sequence 3, Appl	656	5	1.6	20	4	US-08-927-597-80	Sequence 80, Appl
584	5	1.6	14	4	US-09-556-111-3	Sequence 3, Appl	657	5	1.6	20	4	US-08-630-915A-66	Sequence 66, Appl
585	5	1.6	14	4	US-09-556-111-5	Sequence 5, Appl	658	5	1.6	20	4	US-09-304-799-2	Sequence 2, Appl
586	5	1.6	14	4	US-09-316-919-62	Sequence 62, Appl	659	5	1.6	20	4	US-09-556-605-21	Sequence 21, Appl
587	5	1.6	15	1	US-08-685-660A-1	Sequence 1, Appl	660	5	1.6	20	4	US-09-116-492A-8	Sequence 8, Appl
588	5	1.6	15	2	US-08-373-190-4	Sequence 4, Appl	661	5	1.6	21	2	US-09-217-306B-18	Sequence 18, Appl
589	5	1.6	15	2	US-08-974-196-1	Sequence 1, Appl	662	5	1.6	21	2	US-08-470-720-10	Sequence 10, Appl
590	5	1.6	15	2	US-08-476-062A-22	Sequence 22, Appl	663	5	1.6	21	2	US-08-934-915-16	Sequence 16, Appl
591	5	1.6	15	2	US-08-438-190A-4	Sequence 4, Appl	664	5	1.6	21	4	US-08-726-306A-15	Sequence 15, Appl
592	5	1.6	15	3	US-08-350-215-4	Sequence 4, Appl	665	5	1.6	21	4	US-08-895-590-38	Sequence 38, Appl
593	5	1.6	15	3	US-09-287-145A-4	Sequence 4, Appl	666	5	1.6	22	1	US-09-450-972-2	Sequence 2, Appl
594	5	1.6	15	3	US-08-870-529-6	Sequence 6, Appl	667	5	1.6	22	1	US-08-264-002-14	Sequence 14, Appl
595	5	1.6	15	3	US-08-491-954-9	Sequence 9, Appl	668	5	1.6	22	2	US-08-146-028-95	Sequence 95, Appl
596	5	1.6	15	3	US-08-933-983-74	Sequence 74, Appl	669	5	1.6	22	2	US-08-146-028-95	Sequence 95, Appl
597	5	1.6	15	4	US-09-013-896A-29	Sequence 29, Appl	670	5	1.6	22	2	US-08-146-028-96	Sequence 96, Appl
598	5	1.6	15	4	US-08-657-339A-1	Sequence 1, Appl	671	5	1.6	22	2	US-08-146-028-102	Sequence 102, App
599	5	1.6	15	4	US-08-457-694A-1	Sequence 1, Appl	672	5	1.6	22	2	US-08-146-028-103	Sequence 103, App
600	5	1.6	15	4	US-09-556-111-4	Sequence 4, Appl	673	5	1.6	22	4	US-08-723-425A-94	Sequence 94, Appl
601	5	1.6	15	4	US-09-009-953-245	Sequence 245, App	674	5	1.6	22	4	US-08-723-425A-95	Sequence 95, Appl
602	5	1.6	15	4	US-09-009-953-246	Sequence 246, App	675	5	1.6	22	4	US-08-723-425A-96	Sequence 96, Appl
603	5	1.6	15	4	US-09-101-751A-24	Sequence 24, Appl	676	5	1.6	22	4	US-08-723-425A-102	Sequence 102, App
604	5	1.6	15	5	PCT-US96-01314-22	Sequence 22, Appl	677	5	1.6	22	4	US-08-723-425A-103	Sequence 103, App
605	5	1.6	16	1	US-08-333-565-26	Sequence 26, Appl	678	5	1.6	22	4	US-08-723-425A-151	Sequence 151, App
606	5	1.6	16	1	US-08-578-649-13	Sequence 13, Appl	679	5	1.6	22	4	US-09-004-406C-29	Sequence 29, Appl
607	5	1.6	16	2	US-08-661-479-26	Sequence 26, Appl	680	5	1.6	22	4	US-09-112-206-94	Sequence 94, Appl
608	5	1.6	16	3	US-08-822-322-4	Sequence 4, Appl	681	5	1.6	22	4	US-09-112-206-95	Sequence 95, Appl
609	5	1.6	16	4	US-08-602-999A-392	Sequence 392, App	682	5	1.6	22	4	US-09-112-206-96	Sequence 96, Appl
610	5	1.6	16	4	US-09-466-109-4	Sequence 4, Appl	683	5	1.6	22	4	US-09-112-206-102	Sequence 102, App
611	5	1.6	16	4	US-09-227-357-523	Sequence 523, App	684	5	1.6	22	4	US-09-112-206-103	Sequence 103, App

977 5 1.6 80 1 US-08-137-800-41 Sequence 41, App1
978 5 1.6 80 1 US-08-477-383-41 Sequence 41, App1
979 5 1.6 80 1 US-08-487-174-41 Sequence 41, App1
980 5 1.6 80 2 US-08-480-750-41 Sequence 82, App1
981 5 1.6 80 2 US-08-553-501A-85 Sequence 85, App1
982 5 1.6 80 2 US-08-765-783A-104 Sequence 104, App1
983 5 1.6 80 3 US-08-554-840-10 Sequence 10, App1
984 5 1.6 80 3 US-08-554-840-12 Sequence 12, App1
985 5 1.6 80 3 US-08-554-840-13 Sequence 13, App1
986 5 1.6 80 3 US-08-554-840-15 Sequence 15, App1
987 5 1.6 80 3 US-08-894-017-5 Sequence 5, App1
988 5 1.6 80 3 US-09-205-231-85 Sequence 85, App1
989 5 1.6 80 4 US-09-416-557-104 Sequence 104, App1
990 5 1.6 80 4 US-08-818-112-117 Sequence 117, App1
991 5 1.6 80 4 US-08-818-112-117 Sequence 117, App1
992 5 1.6 80 4 US-09-056-556-117 Sequence 117, App1
993 5 1.6 80 4 US-08-925-339-10 Sequence 10, App1
994 5 1.6 80 4 US-08-925-339-12 Sequence 12, App1
995 5 1.6 80 4 US-08-925-339-13 Sequence 13, App1
996 5 1.6 80 4 US-08-925-339-15 Sequence 15, App1
997 5 1.6 80 4 US-09-072-596-112 Sequence 112, App1
998 5 1.6 81 3 US-08-554-840-11 Sequence 11, App1
999 5 1.6 81 4 US-09-564-805-211 Sequence 211, App1
1000 5 1.6 81 4 US-08-856-207A-450 Sequence 450, App1

ALIGNMENTS

RESULT 1
US-09-027-064-2
; Sequence 2, Application US/09027064
; Patent No. 6133006
; GENERAL INFORMATION:
; APPLICANT: SHABON, USMAN
; APPLICANT: BERGSMAN, DEREK
; TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE
; TITLE OF INVENTION: PROTEIN KINASE-HTLAR33
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: 20-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/053,924
; FILING DATE: 28-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-09-027-064-2
Query Match 2.5%; Score 8; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 APVAPPT 186
DB 450 APVAPPT 457

RESULT 2
US-09-271-815-2
; Sequence 2, Application US/09271815
; Patent No. 6297036
; GENERAL INFORMATION:
; APPLICANT: SHABON, USMAN
; APPLICANT: BERGSMAN, DEREK
; TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE PROTEIN KINASE-HTLAR33
; FILE REFERENCE: GH-70172-1
; CURRENT APPLICATION NUMBER: US/09/271,815
; CURRENT FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 09/027,064
; EARLIER FILING DATE: 1998-02-20
; EARLIER APPLICATION NUMBER: 60/053,924
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO: 2
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-271-815-2

Query Match 2.5%; Score 8; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 APVAPPT 186
DB 450 APVAPPT 457

RESULT 3
US-08-258-152-9
; Sequence 9, Application US/08258152
; Patent No. 5686279
; GENERAL INFORMATION:
; APPLICANT: FINER, MITCHELL H.
; APPLICANT: ROBERTS, MARGO R.
; APPLICANT: DULL, THOMAS J.
; APPLICANT: ZSEBO, KRISTINA M.
; APPLICANT: OIN, LU
; TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
; TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CELL GENESYS, INC.
; STREET: 322 LAKESIDE DRIVE
; CITY: FOSTER CITY
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,152
; FILING DATE: 10-JUN-1994
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/076,299
FILING DATE: 11-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I.
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 13.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-358-9600 X131
TELEFAX: 415-349-7392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-258-152-9

Query Match 2.2%; Score 7; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
|||||
Db 2 STSGSGS 8

RESULT 4
US-08-076-299A-9
Sequence 9, Application US/08076299A
Patent No. 5834256
GENERAL INFORMATION:
APPLICANT: FINER, MITCHELL H.
APPLICANT: ROBERTS, MARGO R.
APPLICANT: DULL, THOMAS J.
APPLICANT: ZSEBO, KRISZTINA M.
APPLICANT: QIN, LU
TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
NUMBER OF INVENTION: OF MAMMALIAN CELLS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076,299A
FILING DATE: 11-JUN-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I.
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 13.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-358-9600 X131
TELEFAX: 415-349-7392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-076-299A-9

Query Match 2.2%; Score 7; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
|||||
Db 2 STSGSGS 8

RESULT 5
US-08-438-582-9
Sequence 9, Application US/08438582
Patent No. 5858740
GENERAL INFORMATION:
APPLICANT: FINER, MITCHELL H.
APPLICANT: ROBERTS, MARGO R.
APPLICANT: DULL, THOMAS J.
APPLICANT: ZSEBO, KRISZTINA M.
APPLICANT: QIN, LU
TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
NUMBER OF INVENTION: OF MAMMALIAN CELLS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/438,582
FILING DATE: 10-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,152
FILING DATE: 10-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/076,299
FILING DATE: 11-JUN-93
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I.
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 13.2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-358-9600 X131
TELEFAX: 415-349-7392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-438-582-9

Query Match 2.2%; Score 7; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
|||||
Db 2 STSGSGS 8

```
RESULT 6
US-09-266-596-9
Sequence 9, Application US/09266596
GENERAL INFORMATION:
APPLICANT: FINER, MITCHELL H.
APPLICANT: DULL, THOMAS J.
APPLICANT: ZSEBO, KRISTINA M.
APPLICANT: COOKE, KERRAN
APPLICANT: PARSON, DEBORAH A.
TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESS: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/266,596
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/517,488
FILING DATE: 21-AUG-1995
APPLICATION NUMBER: US 08/258,152
FILING DATE: 10-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/076,299
FILING DATE: 11-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I.
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 13.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-358-9600 X131
TELEFAX: 415-349-7392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-266-596-9

Query Match      2.2%; Score 7; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 41 STSGSGS 47
Db 2 STSGSGS 8

RESULT 7
US-08-479-737-44
Sequence 44, Application US/08479737
GENERAL INFORMATION:
APPLICANT: CAPON, DANIEL J
APPLICANT: WEISS, ARTHUR
APPLICANT: IRVING, BRIAN A
APPLICANT: ROBERTS, MARGO R
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

Zsebo, Kristina
TITLE OF INVENTION: CHIMERIC CHAINS FOR RECEPTOR ASSOCIATED
SIGNAL TRANSDUCTION PATHWAYS
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESS: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

US-08-479-737-44
Sequence 44, Application US/08475442A
GENERAL INFORMATION:
APPLICANT: CAPON, DANIEL J
APPLICANT: WEISS, ARTHUR
APPLICANT: IRVING, BRIAN A
APPLICANT: ROBERTS, MARGO R
APPLICANT: ZSEBO, KRISTINA
TITLE OF INVENTION: CHIMERIC CHAINS FOR
RECEPTOR ASSOCIATED SIGNAL TRANSDUCTION PATHWAYS
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESS: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

US-08-479-737-44
Sequence 44, Application US/08475442A
GENERAL INFORMATION:
APPLICANT: CAPON, DANIEL J
APPLICANT: WEISS, ARTHUR
APPLICANT: IRVING, BRIAN A
APPLICANT: ROBERTS, MARGO R
APPLICANT: ZSEBO, KRISTINA
TITLE OF INVENTION: CHIMERIC CHAINS FOR
RECEPTOR ASSOCIATED SIGNAL TRANSDUCTION PATHWAYS
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESS: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```


SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,442A
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,405
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,194
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/627,643
FILING DATE: 14-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/09431
FILING DATE: 12-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELLS 5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)358-9600x131
TELEFAX: (415)349-7392
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-475-442A-44

Query Match 2.2%; Score 7; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 2 STSGSGS 8

RESULT 9
US-09-029-424-15
Sequence 15, Application US/09029424A
Patent No. 6030795
GENERAL INFORMATION:
APPLICANT: Saitoh, Masao
APPLICANT: Miyazono, Kohel
APPLICANT: Ichijo, Hidenori
TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING TGF RECEPTOR HAVING MODIFIED
TITLE OF INVENTION: GROWTH INHIBITION, AND ITS USE
FILE REFERENCE: L0461/7027
CURRENT APPLICATION NUMBER: US/09/029,424A
CURRENT FILING DATE: 1998-04-28
EARLIER APPLICATION NUMBER: PCT/GB96/02179
EARLIER FILING DATE: 1996-09-04
NUMBER OF SEQ ID NOS: 19
SEQ ID NO 15
LENGTH: 56
TYPE: PRT
ORGANISM: Homo sapiens
US-09-029-424-15

Query Match 2.2%; Score 7; DB 3; Length 56;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 35 STSGSGS 41

RESULT 10
US-09-103-478-19
Sequence 19, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3634
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-19

Query Match 2.2%; Score 7; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 314 YVDPLTV 320
Db 78 YVDPLTV 84

RESULT 11
US-09-193-931C-19
Sequence 19, Application US/09193931C
Patent No. 6320102
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
FILE REFERENCE: 023070-077620
CURRENT APPLICATION NUMBER: US/09/193,931C

CURRENT FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: US 09/103,478
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: US 09/026,221
PRIOR FILING DATE: 1998-02-19
PRIOR APPLICATION NUMBER: US 08/804,534
PRIOR FILING DATE: 1997-02-21
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 19
LENGTH: 90
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (1)..(90)
OTHER INFORMATION: LECT1 HAP3 subunit of CCAAT box-binding factor
OTHER INFORMATION: (CBF) protein B domain homolog
US-09-193-931C-19

Query Match 2.2% Score 7; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
DB 78 YVDPLTV 84

RESULT 12
US-09-103-478-2
Sequence 2, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamara
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberger, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 208 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-103-478-2

Query Match 2.2% Score 7; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
DB 105 YVDPLTV 111

RESULT 13
US-09-193-931C-2
Sequence 2, Application US/09193931C
Patent No. 6320102
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamara
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberger, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: The Regents of the University of California
FILE REFERENCE: 023070-077620
CURRENT APPLICATION NUMBER: US/09/193,931C
CURRENT FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: US 09/103,478
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: US 09/026,221
PRIOR FILING DATE: 1998-02-19
PRIOR APPLICATION NUMBER: US 08/804,534
PRIOR FILING DATE: 1997-02-21
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 2
LENGTH: 208
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-193-931C-2

Query Match 2.2% Score 7; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
DB 105 YVDPLTV 111

RESULT 14
US-08-424-641B-4
Sequence 4, Application US/08424641B
Patent No. 5824523
GENERAL INFORMATION:
APPLICANT: Sylvain Moineau, Shirley A.
APPLICANT: Walker, Ebenezer R. Vedamuthu,
APPLICANT: and Peter A. Vandenberg
TITLE OF INVENTION: Isolated DNA Encoding
TITLE OF INVENTION: Enzyme For Phage
TITLE OF INVENTION: Resistance
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb

MEDIUM TYPE: storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,641B
FILING DATE: April 19, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/366,480
FILING DATE: December 30, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MT 4.1-151
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5824523e
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 Amino Acids
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Peptide
US-08-424-641B-4

Query Match 2.2%; Score 7; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 TVSKIAQ 86
Db 171 TVSKIAQ 177

RESULT 15
US-08-820-980-4
Sequence 4, Application US/08820980
Patent No. 5925388
GENERAL INFORMATION:
APPLICANT: Sylvain Moineau, Shirley A.
APPLICANT: Walker, Ebenezer R. Vedamuthu,
APPLICANT: and Peter A. Vandenberg
TITLE OF INVENTION: Isolated DNA Encoding
TITLE OF INVENTION: Enzyme For Phage
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
MEDIUM TYPE: storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,980
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/424,641
FILING DATE: April 19, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod

REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: Quest 4.1-156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5925388e
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 Amino Acids
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Peptide
US-08-820-980-4

Query Match 2.2%; Score 7; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 TVSKIAQ 86
Db 171 TVSKIAQ 177

RESULT 16
US-08-826-439-4
Sequence 4, Application US/08826439
Patent No. 592673
GENERAL INFORMATION:
APPLICANT: Sylvain Moineau, Shirley A.
APPLICANT: Walker, Ebenezer R. Vedamuthu,
APPLICANT: and Peter A. Vandenberg
TITLE OF INVENTION: Isolated DNA Encoding
TITLE OF INVENTION: Enzyme For Phage
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
MEDIUM TYPE: storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,439
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/424,641
FILING DATE: April 19, 1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: Quest 4.1-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5972673e
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 Amino Acids
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Peptide
US-08-826-439-4

Query Match 2.2%; Score 7; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 TVSKIAQ 86
|||
171 TVSKIAQ 177

DB

RESULT 17
US-08-913-159-8
Sequence 8, Application US/08913159
Patent No. 6500109
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Plasmid-derived type II
restriction-modification systems from *Lactococcus lactis*
NUMBER OF INVENTION: 14
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913.159
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: DK 0179/95
FILING DATE: 17-FEB-1995
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-159-8

Query Match 2.2%; Score 7; DB 4; Length 304;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 TVSKIAQ 86
|||
171 TVSKIAQ 177

DB

RESULT 18
US-09-147-236-11
Sequence 11, Application US/09147236A
Patent No. 6316251
GENERAL INFORMATION:
APPLICANT: TONOUCHI, Naoto
APPLICANT: TSUCHIDA, Takayasu
APPLICANT: YOSHINAGA, Fumihito
APPLICANT: TAHASHI, Naoki
TITLE OF INVENTION: TAKAHISA
TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE
FILE REFERENCE: 6537-011-0PCF
CURRENT APPLICATION NUMBER: US/09/147.236A
EARLIER FILING DATE: 1998-04-08
EARLIER FILING DATE: FCI/JP97/03633
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 11
LENGTH: 344
TYPE: PRT
ORGANISM: *Acetobacter xylinum*
FEATURE:
OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, c, or
US-09-147-236-11

Query Match 2.2%; Score 7; DB 4; Length 344;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 QPAPVP 183
|||
172 QPAPVP 178

DB

RESULT 19
US-08-909-828-1
Sequence 1, Application US/08909828
Patent No. 6060646
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Tolerance of trichothecene Mycotoxins in
Plants and Animals Through the Modification of the
TITLE OF INVENTION: Peptidyl Transferase Gene
NUMBER OF INVENTION: 4
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909.828
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 350 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-909-828-1

Query Match 2.2%; Score 7; DB 3; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TPVVVV 165
|||
81 TPVVVV 87

DB

RESULT 20
US-08-909-828-2
Sequence 2, Application US/08909828
Patent No. 6060646
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Tolerance of Trichothecene Mycotoxins in
Plants and Animals Through the Modification of the
TITLE OF INVENTION: Peptidyl Transferase Gene
NUMBER OF INVENTION: 4
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909.828
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 350 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-909-828-2

Query Match 2.2%; Score 7; DB 3; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 159 TTPVVVV 165
DB 81 TTPVVVV 87

RESULT 21

US-09-232-191-4
Sequence 4, Application US/09232191
Patent No. 6284487
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: Fatty Acid Transport Proteins
FILE REFERENCE: WHI97-21P3ME
CURRENT APPLICATION NUMBER: US/09/232,191
EARLIER APPLICATION NUMBER: 1999-01-14
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-191-4

Query Match 2.2%; Score 7; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 145 TEAPFAT 151

RESULT 22

US-09-232-200-4
Sequence 4, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3ME
CURRENT APPLICATION NUMBER: US/09/232,200A
EARLIER APPLICATION NUMBER: 1999-01-14
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-200-4

Query Match 2.2%; Score 7; DB 4; Length 356;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 145 TEAPFAT 151

RESULT 23

US-09-232-197-4
Sequence 4, Application US/09232197A
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
EARLIER APPLICATION NUMBER: 1999-01-14
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-197-4

Query Match 2.2%; Score 7; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 145 TEAPFAT 151

RESULT 24

US-09-232-201-4
Sequence 4, Application US/09232201A
Patent No. 6348321
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MC
CURRENT APPLICATION NUMBER: US/09/232,201A
EARLIER APPLICATION NUMBER: 1999-01-14
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-201-4

Query Match 2.2%; Score 7; DB 4; Length 356;

Query Match 2.2%; Score 7; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 145 TEAPFAT 151

RESULT 25
US-08-149-105-15
Sequence 15, Application US/08149105
Patent No. 5538892
GENERAL INFORMATION:
APPLICANT: Donahoe, Patricia K.
APPLICANT: Gustafson, Michael
APPLICANT: He, Wei W.
APPLICANT: Wang, Xiao-Fan
TITLE OF INVENTION: TGF- TYPE I RECEPTOR
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,105
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,673
FILING DATE: March 11, 1993
APPLICATION NUMBER: 07/853,396
FILING DATE: March 18, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/211001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 501
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-149-105-15

Query Match 2.2%; Score 7; DB 1; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 187 STSGSGS 193

RESULT 26
US-08-317-847-15
Sequence 15, Application US/08317847
Patent No. 5547854
GENERAL INFORMATION:
APPLICANT: Donahoe, Patricia K.
APPLICANT: Gustafson, Michael

APPLICANT: He, Wei W.
TITLE OF INVENTION: FOUR NOVEL RECEPTORS OF THE TGF-B
TITLE OF INVENTION: FAMILY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,847
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,673
FILING DATE: March 11, 1993
APPLICATION NUMBER: 07/853,396
FILING DATE: March 18, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/127002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 501
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-317-847-15

Query Match 2.2%; Score 7; DB 1; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 187 STSGSGS 193

RESULT 27
US-09-382-256-8
Sequence 8, Application US/09382256A
Patent No. 6207814
GENERAL INFORMATION:
APPLICANT: MIYAZONO, Kohel
TEN DIJKE, Peter
FRANZEN, Petra
YAMASHITA, Hidetoshi
HELDIN, Carl-Henrik
TITLE OF INVENTION: ACTIVITY RECEPTOR LIKE KINASES, PROTEINS
HAVING SERINE THREONINE KINASE DOMAINS,
AND THEIR USE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb

COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/382,256A
FILING DATE: 24-Aug-1999
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: No. 6207814ember 17, 1993
APPLICATION NUMBER: GB 9224057.1
FILING DATE: No. 6207814ember 17, 1992
APPLICATION NUMBER: GB 9304677.9
FILING DATE: March 8, 1993
APPLICATION NUMBER: GB 9304680.3
FILING DATE: March 8, 1993
APPLICATION NUMBER: 9311047.6
FILING DATE: May 28, 1993
APPLICATION NUMBER: 9313763.6
FILING DATE: July 2, 1993
APPLICATION NUMBER: 9316099.2
FILING DATE: August 3, 1993
APPLICATION NUMBER: 321344.5
FILING DATE: October 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: No. 6207814man D. Hanson
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5298.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-382-256-8

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 28
US-09-382-256-16
Sequence 16, Application US/09382256A
Patent No. 6207814
GENERAL INFORMATION:
APPLICANT: MIYAZONO, Kohel
TEN DIJKE, Peter
FRANZEN, Petra
YAMASHITA, Hidetoshi
HELDIN, Carl-Henrik
TITLE OF INVENTION: ACTIVIN RECEPTOR LIKE KINASES, PROTEINS
HAVING SERINE THREONINE KINASE DOMAINS,
AND THEIR USE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/382,256A
FILING DATE: 24-Aug-1999
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: No. 6207814ember 17, 1993
APPLICATION NUMBER: GB 9224057.1
FILING DATE: No. 6207814ember 17, 1992
APPLICATION NUMBER: GB 9304677.9
FILING DATE: March 8, 1993
APPLICATION NUMBER: GB 9304680.3
FILING DATE: March 8, 1993
APPLICATION NUMBER: 9311047.6
FILING DATE: May 28, 1993
APPLICATION NUMBER: 9313763.6
FILING DATE: July 2, 1993
APPLICATION NUMBER: 9316099.2
FILING DATE: August 3, 1993
APPLICATION NUMBER: 321344.5
FILING DATE: October 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: No. 6207814man D. Hanson
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5298.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-382-256-16

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 29
US-09-395-115-8
Sequence 8, Application US/09395115
Patent No. 6271365
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; DiJke, Peter Ten;
FRANZEN, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
HAVING Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/395,115
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6271365ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6271365ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-395-115-8

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 187 STSGSGS 193

RESULT 30
US-09-395-115-16
Sequence 16, Application US/09395115
Patent No. 6271365
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohlei, DiIke, Peter Ten;
APPLICANT: Franzen, Petia; Yamashita, Hidekoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSER: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/395,115

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6271365ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6271365ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-395-115-16

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 187 STSGSGS 193

RESULT 31
US-08-123-934A-8
Sequence 8, Application US/08123934A
Patent No. 6291206
GENERAL INFORMATION:
APPLICANT: WOZNEY, John
APPLICANT: CELESTE, Anthony J.
APPLICANT: THIES, R. Scott
APPLICANT: YAMAJI, No. 6291206oru
TITLE OF INVENTION: RECEPTOR PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute Inc.- Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA USA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,934A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5203
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 876 1170
TELEFAX: 617 876 5851
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-123-934A-8

Query Match 2.2% Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 32

US-08-436-265-8
Sequence 8, Application US/08436265
Patent No. 6316217

GENERAL INFORMATION:

APPLICANT: Miyazono, Kohel; Dijke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/436,265

FILING DATE: 30-October-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/02367

FILING DATE: 17-No. 6316217ember-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9224057.1

FILING DATE: 17-No. 6316217ember-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9304677.9

FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9304680.3

FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9311047.6

FILING DATE: 28-May-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9313763.6

FILING DATE: 2-July-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohel, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-265-8

Query Match 2.2% Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 33

US-08-436-265-16
Sequence 16, Application US/08436265
Patent No. 6316217

GENERAL INFORMATION:

APPLICANT: Miyazono, Kohel; Dijke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/436,265

FILING DATE: 30-October-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/02367

FILING DATE: 17-No. 6316217ember-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9224057.1

FILING DATE: 17-No. 6316217ember-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9304677.9

FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9304680.3

FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9311047.6

FILING DATE: 28-May-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9313763.6

FILING DATE: 2-July-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohnei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-265-16

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
|||
Db 187 STSGSGS 193

RESULT 34

US-09-679-187-8
Sequence 8, Application US/09679187
Patent No. 6331621
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohnei; DiJke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/679,187
FILING DATE: 03-OCT-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-NO. 6331621ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-NO. 6331621ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohnei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-679-187-8

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
|||
Db 187 STSGSGS 193

RESULT 35

US-09-679-187-16
Sequence 16, Application US/09679187
Patent No. 6331621
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohnei; DiJke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/679,187
FILING DATE: 03-OCT-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-NO. 6331621ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-NO. 6331621ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:

FILED DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-679-187-16.

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
|||||
Db 187 STSGSGS 193

RESULT 36
PCT-US94-10080-8
Sequence 8, Application PC/TUS9410080
GENERAL INFORMATION:
APPLICANT: GENETICS INSTITUTE, INC.
TITLE OF INVENTION: RECEPTOR PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute Inc. - Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/10080
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/123,934
FILING DATE: 17-SEP-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5203-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-5851
TELEFAX: (617) 498-8260
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
PCT-US94-10080-8

Query Match 2.2%; Score 7; DB 5; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
|||||
Db 187 STSGSGS 193

RESULT 37
US-08-469-412A-7
Sequence 7, Application US/08469412A
Patent No. 5856125
GENERAL INFORMATION:
APPLICANT: Mavrothalasitis, George J.
APPLICANT: Blair, Donald G.
APPLICANT: Fisher, Robert J.
APPLICANT: Beal Jr., Gregory J.
APPLICANT: Athanasios, Metrop A.
APPLICANT: Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,412A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Garrett-Mackowski, Eugenia
REGISTRATION NUMBER: 37,330
REFERENCE/DOCKET NUMBER: 015280-229000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 543 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..543
OTHER INFORMATION: /note="murine ERF amino acid sequence
OTHER INFORMATION: (first 8 amino acids from first exon not
included)"
US-08-469-412A-7

Query Match 2.2%; Score 7; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 SSGSGLA 57
|||||
Db 398 SSGSGLA 404

RESULT 38

Wed Jul 9 10:00:57 2003

us-10-018-706-2.cligo.ra1

Page 22

US-09-021-715-7
Sequence 7, Application US/09021715
Patent No. 6194547
GENERAL INFORMATION:
APPLICANT: Mavrothalassitis, George J.
Blair, Donald G.
Fisher, Robert J.
Beal Jr., Gregory J.
Athanasios, Meropi A.
Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/021,715
FILING DATE: 10-Feb-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Garselt-Wackowski, Eugenia
REGISTRATION NUMBER: 37,330
REFERENCE/DOCKET NUMBER: 015280-229000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 7:
LENGTH: 543 amino acids
SEQUENCE CHARACTERISTICS:
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1-543
OTHER INFORMATION: /note= "murine ERF amino acid sequence
(filter 8 amino acids from first exon not
included)
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-021-715-7
Query Match 2.2%; Score 7; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 SSGGGIA 57
DB 398 SSGGGIA 404
RESULT 39
US-09-232-200-89
Sequence 89, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MB
CURRENT APPLICATION NUMBER: US/09/232,200A

CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 89
LENGTH: 623
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-200-89
Query Match 2.2%; Score 7; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 186 TEAPFAT 192
DB 398 TEAPFAT 404

RESULT 40
US-09-232-200-98
Sequence 98, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 98
LENGTH: 623
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-200-98
Query Match 2.2%; Score 7; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 186 TEAPFAT 192
DB 398 TEAPFAT 404
Search completed: July 8, 2003, 11:22:01
Job time: 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:21:03 ; Search time 54 Seconds
(without alignments)
694.207 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 322
Sequence: 1 MVTVAINSONQKPIKRLGL.....LFEFRISRMGYVDPITVLK 322

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 445758 seqs, 11641973 residues

Word size: 0

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database:

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PC7_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/PC7US_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.5	223	9	US-09-738-626-3830
2	7	2.2	13	10	US-09-944-411-9
3	7	2.2	13	10	US-09-759-352-44
4	7	2.2	208	9	US-09-533-029-70
5	7	2.2	208	9	US-10-286-264-72
6	7	2.2	213	10	US-10-180-375-195
7	7	2.2	213	10	US-09-771-161A-109
8	7	2.2	251	9	US-10-125-540-300
9	7	2.2	251	10	US-09-764-870-300
10	7	2.2	343	9	US-10-156-761-14239
11	7	2.2	356	10	US-09-943-671-4
12	7	2.2	375	9	US-10-156-761-8098
13	7	2.2	490	9	US-10-156-761-13366
14	7	2.2	496	9	US-10-081-872-212
15	7	2.2	505	10	US-09-771-161A-200
16	7	2.2	505	10	US-09-903-068-8
17	7	2.2	505	10	US-09-903-068-16
18	7	2.2	505	10	US-09-874-628-8
19	7	2.2	806	9	US-09-738-626-5890

20	7	2.2	843	9	US-09-978-303-25	Sequence 25, Appl
21	7	2.2	1336	9	US-10-278-173-128	Sequence 128, Appl
22	7	2.2	1336	9	US-09-934-455-22	Sequence 22, Appl
23	7	2.2	2344	10	US-09-815-242-12713	Sequence 12713, A
24	7	2.2	2796	9	US-09-870-759-114	Sequence 114, Appl
25	7	2.2	3069	9	US-09-712-363-246	Sequence 246, Appl
26	6	1.9	14	9	US-09-956-086-7	Sequence 7, Appl1
27	6	1.9	14	9	US-09-999-745-58	Sequence 58, Appl
28	6	1.9	14	9	US-09-999-745-60	Sequence 60, Appl
29	6	1.9	14	9	US-09-956-087-7	Sequence 7, Appl1
30	6	1.9	14	9	US-09-554-000-42	Sequence 42, Appl
31	6	1.9	14	9	US-09-554-000-44	Sequence 44, Appl
32	6	1.9	14	9	US-09-792-793A-4	Sequence 4, Appl1
33	6	1.9	14	9	US-09-792-793A-6	Sequence 6, Appl1
34	6	1.9	14	9	US-10-158-238-9	Sequence 9, Appl1
35	6	1.9	14	9	US-10-247-279-3	Sequence 3, Appl1
36	6	1.9	14	9	US-10-052-942-6	Sequence 6, Appl1
37	6	1.9	14	10	US-09-883-777-13	Sequence 13, Appl
38	6	1.9	14	10	US-09-987-456-11	Sequence 11, Appl
39	6	1.9	16	10	US-09-883-777-14	Sequence 14, Appl
40	6	1.9	16	10	US-09-883-777-15	Sequence 15, Appl
41	6	1.9	17	9	US-10-091-724-42	Sequence 42, Appl
42	6	1.9	18	9	US-09-956-086-8	Sequence 8, Appl1
43	6	1.9	18	9	US-09-956-086-10	Sequence 10, Appl
44	6	1.9	18	9	US-09-999-745-59	Sequence 59, Appl
45	6	1.9	18	9	US-09-999-745-61	Sequence 61, Appl
46	6	1.9	18	9	US-09-956-087-8	Sequence 8, Appl1
47	6	1.9	18	9	US-09-956-087-10	Sequence 10, Appl
48	6	1.9	18	9	US-09-554-000-43	Sequence 43, Appl
49	6	1.9	18	9	US-09-554-000-45	Sequence 45, Appl
50	6	1.9	18	9	US-09-792-793A-5	Sequence 5, Appl1
51	6	1.9	18	9	US-09-792-793A-7	Sequence 7, Appl1
52	6	1.9	18	9	US-09-866-538-26	Sequence 26, Appl
53	6	1.9	18	9	US-10-158-238-10	Sequence 10, Appl
54	6	1.9	18	9	US-10-158-238-11	Sequence 11, Appl
55	6	1.9	18	9	US-10-121-258-18	Sequence 38, Appl
56	6	1.9	18	9	US-10-247-279-4	Sequence 4, Appl1
57	6	1.9	18	9	US-10-247-279-5	Sequence 5, Appl1
58	6	1.9	19	10	US-09-168-474-19	Sequence 19, Appl
59	6	1.9	22	9	US-09-899-235-20	Sequence 20, Appl
60	6	1.9	22	9	US-10-121-258-19	Sequence 19, Appl
61	6	1.9	22	9	US-10-121-258-19	Sequence 80, Appl
62	6	1.9	28	9	US-10-023-282-800	Sequence 80, Appl
63	6	1.9	32	10	US-09-864-761-38752	Sequence 38752, A
64	6	1.9	33	10	US-09-864-761-38943	Sequence 38943, A
65	6	1.9	34	9	US-10-106-638-8038	Sequence 8038, Appl
66	6	1.9	41	9	US-10-001-876-124	Sequence 124, Appl
67	6	1.9	42	9	US-09-925-299-1393	Sequence 1393, Appl
68	6	1.9	42	10	US-09-925-299-1393	Sequence 799, Appl
69	6	1.9	44	9	US-10-023-282-799	Sequence 799, Appl
70	6	1.9	48	10	US-09-864-761-45369	Sequence 45369, A
71	6	1.9	54	9	US-09-983-802-424	Sequence 224, Appl
72	6	1.9	55	10	US-09-864-761-40446	Sequence 40446, A
73	6	1.9	62	9	US-10-106-638-8547	Sequence 6547, Appl
74	6	1.9	63	9	US-09-738-626-6587	Sequence 6587, Appl
75	6	1.9	65	9	US-10-023-282-449	Sequence 949, Appl
76	6	1.9	69	10	US-09-864-761-45987	Sequence 45987, A
77	6	1.9	76	10	US-09-168-474-6	Sequence 45905, A
78	6	1.9	83	9	US-10-092-154-599	Sequence 599, Appl
79	6	1.9	83	9	US-09-764-847-599	Sequence 599, Appl
80	6	1.9	89	9	US-10-156-761-12606	Sequence 12606, A
81	6	1.9	89	9	US-09-738-626-6090	Sequence 6090, Appl
82	6	1.9	101	9	US-09-867-550-1450	Sequence 1450, Appl
83	6	1.9	104	10	US-09-755-100-7	Sequence 7, Appl1
84	6	1.9	104	10	US-10-023-282-745	Sequence 745, Appl
85	6	1.9	105	10	US-09-864-761-42752	Sequence 42752, A
86	6	1.9	105	10	US-09-925-301-986	Sequence 986, Appl
87	6	1.9	106	10	US-09-734-569-136	Sequence 136, Appl
88	6	1.9	107	10	US-09-740-002-20	Sequence 20, Appl
89	6	1.9	116	9	US-10-073-961-254	Sequence 254, Appl
90	6	1.9	116	10	US-09-764-887-254	Sequence 254, Appl
91	6	1.9	119	10	US-09-820-893-50	Sequence 50, Appl

677	1.9	1426	9	US-10-323-579-15	Sequence 15, Appl	750	1.6	10	9	US-09-572-404B-2432	Sequence 2432, Ap
678	1.9	1541	9	US-09-838-896-3	Sequence 3, Appl	751	1.6	9	US-09-572-404B-2436	Sequence 2436, Ap	
679	1.9	1541	9	US-10-088-805-3	Sequence 3, Appl	752	1.6	10	US-09-572-404B-2461	Sequence 2461, Ap	
680	1.9	1626	9	US-10-188-980-11	Sequence 11, Appl	753	1.6	10	US-09-572-404B-2552	Sequence 2552, Ap	
681	1.9	1681	10	US-09-820-653-3	Sequence 3, Appl	754	1.6	10	US-09-572-404B-2559	Sequence 2559, Ap	
682	1.9	1734	10	US-09-862-927-81	Sequence 81, Appl	755	1.6	10	US-09-572-404B-2621	Sequence 2621, Ap	
683	1.9	1776	10	US-09-841-132-179	Sequence 179, Appl	756	1.6	10	US-09-572-404B-2622	Sequence 2622, Ap	
684	1.9	1848	9	US-09-833-996-6	Sequence 6, Appl	757	1.6	10	US-09-572-404B-2660	Sequence 2660, Ap	
685	1.9	1848	9	US-10-080-505-6	Sequence 6, Appl	758	1.6	10	US-09-572-404B-2694	Sequence 2694, Ap	
686	1.9	2061	9	US-10-060-036-181	Sequence 181, Appl	759	1.6	10	US-09-572-404B-2894	Sequence 2894, Ap	
687	1.9	2364	9	US-10-156-761-7834	Sequence 7834, Ap	760	1.6	10	US-09-572-404B-3224	Sequence 3224, Ap	
688	1.9	2402	9	US-10-203-224-2	Sequence 2, Appl	761	1.6	10	US-09-572-404B-3270	Sequence 3270, Ap	
689	1.9	2476	9	US-09-824-574-7	Sequence 7, Appl	762	1.6	10	US-09-572-404B-3272	Sequence 3272, Ap	
690	1.9	2502	10	US-09-772-316-1	Sequence 1, Appl	763	1.6	10	US-09-572-404B-3283	Sequence 3283, Ap	
691	1.9	2665	10	US-09-864-761-34248	Sequence 34248, A	764	1.6	10	US-09-572-404B-3283	Sequence 3290, Ap	
692	1.9	2749	9	US-10-227-338-4	Sequence 4, Appl	765	1.6	10	US-09-572-404B-3282	Sequence 3292, Ap	
693	1.9	3353	10	US-09-888-615-64	Sequence 64, Appl	766	1.6	10	US-09-572-404B-3284	Sequence 3296, Ap	
694	1.9	3571	10	US-10-150-821-2	Sequence 2, Appl	767	1.6	10	US-09-572-404B-3286	Sequence 3296, Ap	
695	1.9	3571	10	US-09-911-842-2	Sequence 2, Appl	768	1.6	10	US-09-572-404B-3450	Sequence 3296, Ap	
696	1.9	3594	9	US-10-150-821-4	Sequence 4, Appl	769	1.6	10	US-09-572-404B-3450	Sequence 3296, Ap	
697	1.9	3594	10	US-09-911-842-4	Sequence 4, Appl	770	1.6	10	US-09-572-404B-3528	Sequence 3296, Ap	
698	1.9	3664	9	US-10-156-761-10435	Sequence 10435, A	771	1.6	10	US-09-572-404B-3528	Sequence 3328, Ap	
699	1.9	4840	9	US-10-177-293-423	Sequence 423, Appl	772	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
700	1.6	5	9	US-10-093-200A-12	Sequence 6, Appl	773	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
701	1.6	7	9	US-10-093-200A-12	Sequence 12, Appl	774	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
702	1.6	7	9	US-10-093-200A-12	Sequence 12, Appl	775	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
703	1.6	8	9	US-09-765-555-61	Sequence 61, Appl	776	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
704	1.6	8	9	US-09-813-153-237	Sequence 237, Appl	777	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
705	1.6	8	9	US-10-104-919-48	Sequence 48, Appl	778	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
706	1.6	9	10	US-09-780-668A-24	Sequence 24, Appl	779	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
707	1.6	10	9	US-10-040-572-4A-152	Sequence 152, Appl	780	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
708	1.6	10	9	US-09-572-404B-452	Sequence 452, Appl	781	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
709	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	782	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
710	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	783	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
711	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	784	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
712	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	785	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
713	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	786	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
714	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	787	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
715	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	788	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
716	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	789	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
717	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	790	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
718	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	791	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
719	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	792	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
720	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	793	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
721	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	794	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
722	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	795	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
723	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	796	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
724	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	797	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
725	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	798	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
726	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	799	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
727	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	800	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
728	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	801	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
729	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	802	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
730	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	803	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
731	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	804	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
732	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	805	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
733	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	806	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
734	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	807	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
735	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	808	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
736	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	809	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
737	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	810	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
738	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	811	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
739	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	812	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
740	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	813	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
741	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	814	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
742	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	815	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
743	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	816	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
744	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	817	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
745	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	818	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
746	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	819	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
747	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	820	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
748	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	821	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	
749	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	822	1.6	10	US-09-572-404B-3532	Sequence 3328, Ap	

969 5 1.6 26 9 US-09-969-730-272 Sequence 272, App
970 5 1.6 26 9 US-10-023-282-1220 Sequence 1220, App
971 5 1.6 26 9 US-10-213-512-177 Sequence 1120, App
972 5 1.6 26 10 US-09-864-761-40782 Sequence 40782, A
973 5 1.6 26 10 US-09-864-761-41159 Sequence 41159, A
974 5 1.6 26 10 US-09-864-761-43426 Sequence 43426, A
975 5 1.6 26 10 US-09-921-397-7 Sequence 7, App1
976 5 1.6 26 10 US-09-921-397-35 Sequence 35, App1
977 5 1.6 26 10 US-09-071-838-17 Sequence 17, App1
978 5 1.6 27 9 US-09-745-078A-11 Sequence 11, App1
979 5 1.6 27 9 US-10-062-548-165 Sequence 165, App
980 5 1.6 27 10 US-09-864-761-40237 Sequence 40237, A
981 5 1.6 27 10 US-09-864-761-46273 Sequence 46273, A
982 5 1.6 27 10 US-09-939-980-531 Sequence 531, App
983 5 1.6 27 10 US-09-825-012-30 Sequence 30, App1
984 5 1.6 28 9 US-09-999-724-28 Sequence 28, App1
985 5 1.6 28 9 US-09-745-078A-10 Sequence 10, App1
986 5 1.6 28 10 US-09-921-397-19 Sequence 19, App1
987 5 1.6 29 9 US-09-745-078A-9 Sequence 9, App1
988 5 1.6 29 9 US-10-091-504-674 Sequence 674, App
989 5 1.6 29 10 US-09-864-761-48468 Sequence 48468, A
990 5 1.6 29 10 US-09-764-869-674 Sequence 674, App
991 5 1.6 29 10 US-09-935-291A-27 Sequence 27, App1
992 5 1.6 29 10 US-09-935-291A-28 Sequence 28, App1
993 5 1.6 29 10 US-09-935-291A-29 Sequence 29, App1
994 5 1.6 29 10 US-09-935-291A-30 Sequence 30, App1
995 5 1.6 29 10 US-08-335-291A-31 Sequence 31, App1
996 5 1.6 29 10 US-08-335-291A-32 Sequence 32, App1
997 5 1.6 29 10 US-08-335-291A-33 Sequence 33, App1
998 5 1.6 29 10 US-08-935-291A-34 Sequence 34, App1
999 5 1.6 29 10 US-09-902-741-5 Sequence 5, App1
1000 5 1.6 30 9 US-09-745-078A-8 Sequence 8, App1

ALIGNMENTS

RESULT 1
US-09-738-626-3830
; Sequence 3830, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO: 3830
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3830
Query Match 2.5%; Score 8; DB 9; Length 223;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 HTTPSVAV 142
Db 130 HTTPSVAV 137

RESULT 2
US-09-944-411-9
; Sequence 9, Application US/09944411
; Patent No. US20020106799A1
; GENERAL INFORMATION:
; APPLICANT: FINER, MITCHELL H.
; DULU, THOMAS J.
; ZSEHO, KRISTINA M.
; COOKE, KERRAN
; FARSON, DEBORAH A.
; TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
; VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
; OF MAMMALIAN CELLS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESSES:
; ADDRESSER: CELL GENESYS, INC.
; STREET: 322 LAKEVIEW DRIVE
; CITY: FOSTER CITY
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/944,411
; FILING DATE: 04-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/914,893
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/258,152
; FILING DATE: 10-JUN-1998
; APPLICATION NUMBER: US 08/076,299
; FILING DATE: 11-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KRUPEN, KAREN I.
; REGISTRATION NUMBER: 34,647
; REFERENCE/DOCKET NUMBER: CELL 13.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-358-9600 X131
; TELEFAX: 415-349-7392
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: peptide
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-944-411-9
Query Match 2.2%; Score 7; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
Db 2 STSGSGS 8
RESULT 3
US-09-759-352-44
; Sequence 44, Application US/09759352
; Patent No. US2002011474A1

GENERAL INFORMATION:
APPLICANT: CAPON, DANIEL J
WEISS, ARTHUR
IRVING, BRYAN A
ROBERTS, MARGO R
ZSEBO, KRISZTINA
TITLE OF INVENTION: CHIMERIC CHAINS FOR RECEPTOR-ASSOCIATED
SIGNAL TRANSDUCTION PATHWAYS
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKEVIEW DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/759,352
FILING DATE: 16-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/567,393
FILING DATE: 01-DEC-1995
APPLICATION NUMBER: US 08/475,442
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/238,405
FILING DATE: 05-MAY-1994
APPLICATION NUMBER: US 07/988,194
FILING DATE: 09-DEC-1992
APPLICATION NUMBER: US 07/627,643
FILING DATE: 13-DEC-1990
APPLICATION NUMBER: WO PCT/US91/09431
FILING DATE: 12-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 5.13
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)358-9600 X131
TELEFAX: (415)349-7392
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-759-352-44
Query Match 2.2%; Score 7; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
Db 2 STSGSGS 8
RESULT 4
US-09-533-029-70
Sequence 70, Application US/09533029
Publication No. US20030046723A1
GENERAL INFORMATION:
APPLICANT: Heard, Jacqueline
APPLICANT: Broun, Pierre
APPLICANT: Riechmann, Jose-Luis
APPLICANT: Keddie, James

APPLICANT: Pineda, Omaira
APPLICANT: Adam, Luc
APPLICANT: Samaha, Raymond
APPLICANT: Zhang, James
APPLICANT: Yu, Guo-Liang
APPLICANT: Ratcliffe, Oliver
APPLICANT: Pilgrim, Marsha
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Reuber, Lynne
TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
FILE REFERENCE: MBI-010
CURRENT APPLICATION NUMBER: US/09/533,029
CURRENT FILING DATE: 2000-03-22
EARLIER APPLICATION NUMBER: 60/125,814
EARLIER FILING DATE: 1999-03-23
NUMBER OF SEQ ID NOS: 121
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 70
LENGTH: 208
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G620
US-09-533-029-70
Query Match 2.2%; Score 7; DB 9; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 YVDPLTV 320
Db 105 YVDPLTV 111
RESULT 5
US-10-286-264-72
Sequence 72, Application US/10286264
Publication No. US20030093837A1
GENERAL INFORMATION:
APPLICANT: Keddie, James
APPLICANT: Riechmann, Jose-Luis
APPLICANT: Ratcliffe, Oliver
APPLICANT: Zhang, James
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Pineda, Omaira
APPLICANT: Yu, Guo-Liang
APPLICANT: Heard, Jacqueline
APPLICANT: Adam, Luc
APPLICANT: Broun, Pierre
APPLICANT: Reuber, Lynne
APPLICANT: Pilgrim, Marsha
APPLICANT: Samaha, Raymond
TITLE OF INVENTION: POLYNUCLEOTIDES FOR SEED TRAIT ALTERATION
FILE REFERENCE: MBI-008
CURRENT APPLICATION NUMBER: US/10/286,264
CURRENT FILING DATE: 2002-11-01
PRIOR APPLICATION NUMBER: 60/125,814
PRIOR FILING DATE: 1999-03-23
NUMBER OF SEQ ID NOS: 165
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 72
LENGTH: 208
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G620
US-10-286-264-72
Query Match 2.2%; Score 7; DB 9; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 YVDPLTV 320

Db 105 YVDP LTV 111

RESULT 6
US-10-180-375-195
Sequence 195, Application US/10180375
Publication No. US20030126638A1
GENERAL INFORMATION:
APPLICANT: Allen, William B.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Famodu, Omolayo O.
APPLICANT: Harvell, Leslie T.
APPLICANT: Helentjaris, Timothy
APPLICANT: Li, Changjiang
APPLICANT: Lowe, Keith
APPLICANT: Oliveira, Igor Cunha
APPLICANT: Shen, Bo
APPLICANT: Tarczyński, Mitchell C.
TITLE OF INVENTION: Alteration of Oil Traits in Plants
FILE REFERENCE: B01458 US N01
CURRENT APPLICATION NUMBER: US/10/180,375
CURRENT FILING DATE: 2002-06-26
NUMBER OF SEQ ID NOS: 222
SOFTWARE: Microsoft Office 97
SEQ ID NO 195
LENGTH: 208
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION:
OTHER INFORMATION: gi 6552738
US-10-180-375-195

Query Match 2.2%; Score 7; DB 9; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 YVDP LTV 320
Db 105 YVDP LTV 111

RESULT 7
US-09-771-161A-109
Sequence 109, Application US/09771161A
Patent No. US20020110811A1
GENERAL INFORMATION:
APPLICANT: LEVINE, et al.
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
FILE REFERENCE: 802620-2005.1
CURRENT APPLICATION NUMBER: US/09/771,161A
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 09/724,676
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: 135776
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 135619
PRIOR FILING DATE: 2000-04-12
NUMBER OF SEQ ID NOS: 273
SOFTWARE: PatentIn version 3.0
SEQ ID NO 109
LENGTH: 213
TYPE: PRT
ORGANISM: Homo sapiens
US-09-771-161A-109

Query Match 2.2%; Score 7; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47

Db 187 STSGSGS 193

RESULT 8
US-10-125-540-300
Sequence 300, Application US/10125540
Publication No. US20030059875A1
GENERAL INFORMATION:
APPLICANT: Noren, et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P0214C1
CURRENT APPLICATION NUMBER: US/10/125,540
CURRENT FILING DATE: 2002-04-19
Prior Application removed - See file wrapper or Palm
NUMBER OF SEQ ID NOS: 646
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 300
LENGTH: 251
TYPE: PRT
ORGANISM: Homo sapiens
US-10-125-540-300

Query Match 2.2%; Score 7; DB 9; Length 251;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 KPPTPP 173
Db 211 KPPTPP 217

RESULT 9
US-09-764-870-300
Sequence 300, Application US/09764870
Patent No. US20020042386A1
GENERAL INFORMATION:
APPLICANT: Rosen, et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P0214
CURRENT APPLICATION NUMBER: US/09/764,870
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 646
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 300
LENGTH: 251
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-870-300

Query Match 2.2%; Score 7; DB 10; Length 251;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 KPPTPP 173
Db 211 KPPTPP 217

RESULT 10
US-10-156-761-14239
Sequence 14239, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMTURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14239
LENGTH: 343
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14239

Query Match 2.2%; Score 7; DB 9; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GTATVAG 221
Db 63 GTATVAG 69

RESULT 11
US-09-943-671-4
Sequence 4, Application US/09943671
Patent No. US2002010673A1
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: Fatty Acid Transport Proteins
FILE REFERENCE: WHI97-21p3ME
CURRENT APPLICATION NUMBER: US/09/943,671
CURRENT FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 09/232,191
PRIOR FILING DATE: 1999-10-14
PRIOR APPLICATION NUMBER: 60/093,491
PRIOR FILING DATE: 1998-07-20
PRIOR APPLICATION NUMBER: 60/110,941
PRIOR FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-943-671-4

Query Match 2.2%; Score 7; DB 10; Length 356;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
Db 145 TEAPFAT 151

RESULT 12
US-10-156-761-8098
Sequence 8098, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8098
LENGTH: 375
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8098

Query Match 2.2%; Score 7; DB 9; Length 375;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GTATVAG 221
Db 176 GTATVAG 182

RESULT 13
US-10-156-761-13366
Sequence 13366, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13366
LENGTH: 490
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13366

Query Match 2.2%; Score 7; DB 9; Length 490;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 APVAPV 185
Db 38 APVAPV 44

RESULT 14
US-10-081-872-212
Sequence 212, Application US/10081872
Publication No. US20030125534A1
GENERAL INFORMATION:
APPLICANT: Callen, Walter
APPLICANT: Richardson, Toby
APPLICANT: Frey, Gerhard
APPLICANT: Short, Jay M.
APPLICANT: Machur, Eric J.
APPLICANT: Gray, Kevin A.
APPLICANT: Kerovuo, Janne S.
APPLICANT: Slupeska, Malgorzata
TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
FILE REFERENCE: 09010-108001

CURRENT APPLICATION NUMBER: US/10/081,872
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: US 60/270,495
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: US 60/270,496
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: US 60/291,122
PRIOR FILING DATE: 2001-05-14
NUMBER OF SEQ ID NOS: 321
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 212
LENGTH: 496
TYPE: PRT
ORGANISM: Environmental
US-10-081-872-212

Query Match 2.2%; Score 7; DB 9; Length 496;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 NNNSSY 105
Db 167 NNNSSY 173

RESULT 15
US-09-771-161A-200
Sequence 200, Application US/09771161A
Patent No. US20020110811A
GENERAL INFORMATION:
APPLICANT: LEVINE, et al
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
FILE REFERENCE: 802620-2005.1
CURRENT APPLICATION NUMBER: US/09/771,161A
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 09/724,676
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: 136776
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 135619
PRIOR FILING DATE: 2000-04-12
NUMBER OF SEQ ID NOS: 273
SOFTWARE: PatentIn version 3.0
SEQ ID NO 200
LENGTH: 505
TYPE: PRT
ORGANISM: Homo sapiens
US-09-771-161A-200

Query Match 2.2%; Score 7; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 16
US-09-903-068-8
Sequence 8, Application US/09903068
Patent No. US20020123139A1
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; DiJke, Peter Ten;
Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York

ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/903,068
FILING DATE: 11-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/679,187
FILING DATE: <unknown>
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. US20020123139A1ember-1993
APPLICATION NUMBER: 9234057
FILING DATE: 17-No. US20020123139A1ember-1992
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohel, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 638-3884

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: Protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-903-068-8

Query Match 2.2%; Score 7; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 17
US-09-903-068-16
Sequence 16, Application US/09903068
Patent No. US20020123139A1
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; DiJke, Peter Ten;
Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/903,068
FILING DATE: 11-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/679,187
FILING DATE: <Unknown>
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. US20020123139A1ember-1993
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. US20020123139A1ember-1992
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 638-3884
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-903-068-16

Query Match 2.2%; Score 7; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 18
US-09-874-628-8
Sequence 8, Application US/09874628
Patent No. US20020137133A1
GENERAL INFORMATION:
APPLICANT: WOZNEY, John
CELESTE, Anthony J.
THIRS, R. Scott
YAMAJI, No. US20020137133A1
TITLE OF INVENTION: RECEPTOR PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute Inc.- Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/874,628

FILING DATE: 05-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/123,934
FILING DATE: 17-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5203
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 876 1170
TELEFAX: 617 876 5851
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-874-628-8

Query Match 2.2%; Score 7; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 19
US-09-738-626-5890
Sequence 5890, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentin ver. 3.0
SEQ ID NO 5890
LENGTH: 806
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5890

Query Match 2.2%; Score 7; DB 9; Length 806;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 230 WFSGRDG 236
Db 631 WFSGRDG 637

RESULT 20

```
US-09-978-303-25
; Sequence 25, Application US/09978303
; Publication No. US20030049728A1
; GENERAL INFORMATION:
; APPLICANT: Julius, David J.
; APPLICANT: Caterina, Michael J.
; APPLICANT: Blake, Anthony J.
; TITLE OF INVENTION: Nucleic acid sequences encoding
; TITLE OF INVENTION: capsacin receptor and capsacin receptor-related
; FILE REFERENCE: UCA1084CON
; CURRENT APPLICATION NUMBER: US/09/978,303
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/235,451
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/072,151
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 08/915,461
; PRIOR FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 843
; TYPE: PRT
; ORGANISM: chicken
US-09-978-303-25
```

```
Query Match      2.2%; Score 7; DB 9; Length 843;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      80 TVSKIAQ 86
Db      692 TVSKIAQ 698
```

```
RESULT 21
US-10-278-173-128
; Sequence 128, Application US/10278173
; Publication No. US20030061637A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Brown, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Pineda, Omatra
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Keddle, James
; APPLICANT: Heard, Jacqueline
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
; FILE REFERENCE: MBI-009
; CURRENT APPLICATION NUMBER: US/10/278,173
; PRIOR FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US/09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 128
; LENGTH: 1336
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; OTHER INFORMATION: G987
US-10-278-173-128
```

```
Query Match      2.2%; Score 7; DB 9; Length 1336;
```

```
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      61 QVITDSQ 67
Db      779 QVITDSQ 785
```

```
RESULT 22
US-09-934-455-22
; Sequence 22, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omatra
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV
; FILE REFERENCE: MBI-0025
; CURRENT APPLICATION NUMBER: US/09/934,455
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/227439
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: MBI-0022
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: MBI-0023
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 516
; SOFTWARE: Patent version 3.1
; SEQ ID NO 22
; LENGTH: 1336
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-934-455-22
```

```
Query Match      2.2%; Score 7; DB 9; Length 1336;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      61 QVITDSQ 67
Db      779 QVITDSQ 785
```

```
RESULT 23
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlisen, Kari L.
; APPLICANT: Zyking, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Rawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 12713
;; LENGTH: 2344
;; TYPE: PRT
;; ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 2.2%; Score 7; DB 10; Length 2344;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STGSGS 47
DB 2088 STGSGS 2094

RESULT 24
US-09-870-759-114
;; Sequence 114, Application US/09870759
;; Patent No. US20020177551A1
;; GENERAL INFORMATION:
;; APPLICANT: TERMAN, David S
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
;; FILE REFERENCE: 870759
;; CURRENT APPLICATION NUMBER: US/09/870,759
;; PRIOR FILING DATE: 2002-01-14
;; PRIOR APPLICATION NUMBER: US 60/208,128
;; PRIOR FILING DATE: 2000-05-30
;; NUMBER OF SEQ ID NOS: 166
;; SOFTWARE: Patent version 3.1
;; SEQ ID NO: 114
;; LENGTH: 2796
;; TYPE: PRT
;; ORGANISM: Mycobacterium bovis
US-09-870-759-114

Query Match 2.2%; Score 7; DB 9; Length 2796;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 IHKDAQ 278
DB 2088 IHKDAQ 2094

RESULT 25
US-09-712-363-246
;; Sequence 246, Application US/09712363
;; Patent No. US20020164588A1
;; GENERAL INFORMATION:
;; APPLICANT: Eisenberg, David
;; APPLICANT: Rotstein, Sergio H.
;; APPLICANT: Marcotte, Edward M.
;; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;; FILE REFERENCE: 07419-032001
;; CURRENT APPLICATION NUMBER: US/09/712,363
;; PRIOR FILING DATE: 2000-11-13
;; PRIOR APPLICATION NUMBER: PCT/US00/02246
;; PRIOR FILING DATE: 2000-01-28
;; PRIOR APPLICATION NUMBER: 60/179,531

;; PRIOR FILING DATE: 2000-02-01
;; PRIOR APPLICATION NUMBER: 60/117,844
;; PRIOR FILING DATE: 1999-01-29
;; PRIOR APPLICATION NUMBER: 60/118,206
;; PRIOR FILING DATE: 1999-02-01
;; PRIOR APPLICATION NUMBER: 60/126,593
;; PRIOR FILING DATE: 1999-03-26
;; PRIOR APPLICATION NUMBER: 60/134,093
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 60/134,092
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 60/165,124
;; PRIOR FILING DATE: 1999-11-12
;; PRIOR APPLICATION NUMBER: 60/165,086
;; PRIOR FILING DATE: 1999-11-12
;; NUMBER OF SEQ ID NOS: 292
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 246
;; LENGTH: 3069
;; TYPE: PRT
;; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-246

Query Match 2.2%; Score 7; DB 9; Length 3069;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 IHKDAQ 278
DB 2206 IHKDAQ 2212

RESULT 26
US-09-956-086-7
;; Sequence 7, Application US/09956086
;; Patent No. US2002015498A1
;; GENERAL INFORMATION:
;; APPLICANT: FILPULA, DAVID
;; APPLICANT: WANG, MAOLIANG
;; APPLICANT: SHORR, ROBERT
;; APPLICANT: WHITLOW, MARC
;; APPLICANT: LEE, LILSYNG S.
;; TITLE OF INVENTION: SINGLE-CHAIN ANTIGEN-BINDING PROTEINS
;; CAPABLE OF GLYCOSYLATION, PRODUCTION AND USES THEREOF
;; NUMBER OF SEQUENCES: 33
;; CORRESPONDENCE ADDRESSES:
;; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;; STREET: 1100 NEW YORK AVE., NW, SUITE 600
;; CITY: WASHINGTON
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/956,086
;; FILING DATE: 20-Sep-2001
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/069,821
;; FILING DATE: <Unknown>
;; APPLICATION NUMBER: US 60/063,074
;; FILING DATE: 27-OCT-1997
;; APPLICATION NUMBER: US 60/050,472
;; FILING DATE: 23-JUN-1997
;; APPLICATION NUMBER: US 60/044,449
;; FILING DATE: 30-APR-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KIM, JUDITH U.
;; REGISTRATION NUMBER: 40,679

```

; REFERENCE/DOCKET NUMBER: 0977.2280003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 14 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: No. US20020155498A1 Relevant
;     MOLECULE TYPE: Peptide
;     SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-956-086-7

Query Match      1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      41 STSGSG 46
        |||||
        2 STSGSG 7

Db

RESULT 27
US-09-999-745-58
; Sequence 58, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patent version 3.0
; SEQ ID NO 58
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Linker moiety
US-09-999-745-58

Query Match      1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      41 STSGSG 46
        |||||
        2 STSGSG 7

Db

RESULT 28
US-09-999-745-60
; Sequence 60, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patent version 3.0
; SEQ ID NO 60
```

```

; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Linker moiety
US-09-999-745-60

Query Match      1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      41 STSGSG 46
        |||||
        2 STSGSG 7

Db

RESULT 29
US-09-956-087-7
; Sequence 7, Application US/09956087
; Patent No. US20020161201A1
; GENERAL INFORMATION:
; APPLICANT: FILIPULA, DAVID
; APPLICANT: WANG, MAOLIANG
; APPLICANT: SHORR, ROBERT
; APPLICANT: WHITLOW, MARC
; APPLICANT: LEE, LHSYNG S.
; TITLE OF INVENTION: SINGLE-CHAIN ANTIGEN-BINDING PROTEINS
; CAPABLE OF GLYCOSYLATION, PRODUCTION AND USES THEREOF
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSER: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/956,087
; FILING DATE: 20-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/069,821
; FILING DATE: 1998-04-30
; APPLICATION NUMBER: US 60/063,074
; FILING DATE: 27-OCT-1997
; APPLICATION NUMBER: US 60/050,472
; FILING DATE: 23-JUN-1997
; APPLICATION NUMBER: US 60/044,449
; FILING DATE: 30-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KIM, JUDITH U
; REGISTRATION NUMBER: 40,679
; REFERENCE/DOCKET NUMBER: 0977.2280003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 14 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: not relevant
;     MOLECULE TYPE: peptide
;     SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-956-087-7

Query Match      1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
```

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 30
US-09-554-000-42
; Sequence 42, Application US/09554000
; Patent No. US20020165364A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: DETECTION OF ANALYTES
; FILE REFERENCE: 07257/042001
; CURRENT APPLICATION NUMBER: US/09/554,000
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 08/818,252
; PRIOR FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Linker moiety
US-09-554-000-42

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 31
US-09-554-000-44
; Sequence 44, Application US/09554000
; Patent No. US20020165364A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: DETECTION OF ANALYTES
; FILE REFERENCE: 07257/042001
; CURRENT APPLICATION NUMBER: US/09/554,000
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 08/818,252
; PRIOR FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Linker moiety
US-09-554-000-44

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 32
US-09-792-793A-4
; Sequence 4, Application US/09792793A
; Patent No. US20020168370A1
; GENERAL INFORMATION:
; APPLICANT: McDonald, John R.
; APPLICANT: Cogging, Philip
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING SECONDARY TISSUE DAMAGE AND
; TITLE OF INVENTION: OTHER INFLAMMATORY CONDITIONS AND DISORDERS
; FILE REFERENCE: 25020-601D
; CURRENT APPLICATION NUMBER: US/09/792,793A
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 14
; TYPE: PRT
; ORGANISM: homo sapien
; FEATURE:
; OTHER INFORMATION: peptide linker for use in conjugates
US-09-792-793A-4

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 33
US-09-792-793A-6
; Sequence 6, Application US/09792793A
; Patent No. US20020168370A1
; GENERAL INFORMATION:
; APPLICANT: McDonald, John R.
; APPLICANT: Cogging, Philip
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING SECONDARY TISSUE DAMAGE AND
; TITLE OF INVENTION: OTHER INFLAMMATORY CONDITIONS AND DISORDERS
; FILE REFERENCE: 25020-601D
; CURRENT APPLICATION NUMBER: US/09/792,793A
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 14
; TYPE: PRT
; ORGANISM: homo sapien
; FEATURE:
; OTHER INFORMATION: peptide linker for use in conjugates
US-09-792-793A-6

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 34
US-10-158-238-9
; Sequence 9, Application US/10158238
; Publication No. US20030040604A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Marken, John S.
; APPLICANT: Marken, John S.
; TITLE OF INVENTION: STGLC-12 POLYPEPTIDES, POLYNUCLEOTIDES, AND METHODS OF USE THERE
; FILE REFERENCE: 3290-A
; CURRENT APPLICATION NUMBER: US/10/158,238

CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: US 60/294,199
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Peptide linker sequence
US-10-158-238-9

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 35
US-10-247-279-3
Sequence 3, Application US/10247279
Publication No. US20030086928A1
GENERAL INFORMATION:
APPLICANT: Montminy, Marc
TITLE OF INVENTION: COMPOSITIONS ASSOCIATED WITH COMPLEX
FILE REFERENCE: SALKINS.036CP1
CURRENT FILING DATE: 2002-09-17
PRIOR APPLICATION NUMBER: US/10/247,279
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/190,705
PRIOR FILING DATE: 2001-03-19
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Linker
US-10-247-279-3

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 36
US-10-052-942-6
Sequence 6, Application US/10052942
Publication No. US20030104402A1
GENERAL INFORMATION:
APPLICANT: Zauderer, Maurice
APPLICANT: Smith, Ernest
TITLE OF INVENTION: Methods of Producing or Identifying Intrabodies in Eukaryotic Cell
FILE REFERENCE: 1821.0090004
CURRENT APPLICATION NUMBER: US/10/052,942
PRIOR FILING DATE: 2002-01-23
PRIOR APPLICATION NUMBER: 60/298,095
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/271,422
PRIOR FILING DATE: 2001-02-27

PRIOR APPLICATION NUMBER: 60/263,200
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,225
PRIOR FILING DATE: 2001-01-23
NUMBER OF SEQ ID NOS: 154
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Linker
US-10-052-942-6

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 37
US-09-883-777-13
Sequence 13, Application US/09883777
Patent No. US20020110853A1
GENERAL INFORMATION:
APPLICANT: Wiley, Steven R.
TITLE OF INVENTION: TWEAK RECEPTOR
FILE REFERENCE: 2868-C
CURRENT APPLICATION NUMBER: US/09/883,777
CURRENT FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: US 60/172,878
PRIOR FILING DATE: 1998-12-20
PRIOR APPLICATION NUMBER: US 60/203,347
PRIOR FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: PCT/US00/34755
PRIOR FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: US 09/742,454
PRIOR FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Linker Moleley
US-09-883-777-13

Query Match 1.9%; Score 6; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 38
US-09-987-456-11
Sequence 11, Application US/09987456
Patent No. US20020123057A1
GENERAL INFORMATION:
APPLICANT: University of Rochester
APPLICANT: Zauderer, Maurice
APPLICANT: Ernest S. Smith
TITLE OF INVENTION: In Vitro Methods of Producing and Selecting
FILE REFERENCE: 1821.0070004
CURRENT APPLICATION NUMBER: US/09/987,456
CURRENT FILING DATE: 2001-11-14

; PRIOR APPLICATION NUMBER: 60/271,424
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/262,067
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/298,087
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/249,268
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: linker peptide
US-09-987-456-11

Query Match 1.9%; Score 6; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 39
US-09-883-777-14
; Sequence 14, Application US/09883777
; Patent No. US20020110853A1
; GENERAL INFORMATION:
; APPLICANT: Wiley, Steven R.
; TITLE OF INVENTION: TWEAK RECEPTOR
; FILE REFERENCE: 2968-C
; CURRENT APPLICATION NUMBER: US/09/883,777
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/172,878
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: US 60/203,347
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: PCT/US00/34755
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/742,454
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Linker Moiety
US-09-883-777-14

Query Match 1.9%; Score 6; DB 10; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 40
US-09-883-777-15
; Sequence 15, Application US/09883777
; Patent No. US20020110853A1
; GENERAL INFORMATION:
; APPLICANT: Wiley, Steven R.
; TITLE OF INVENTION: TWEAK RECEPTOR
; FILE REFERENCE: 2968-C
; CURRENT APPLICATION NUMBER: US/09/883,777

; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/172,878
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: US 60/203,347
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: PCT/US00/34755
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/742,454
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Linker Moiety
US-09-883-777-15

Query Match 1.9%; Score 6; DB 10; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

Search completed: July 8, 2003, 11:30:23
Job time : 74 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:16:18 ; Search time 40 Seconds
(without alignments)
773.882 Million cell updates/sec

Title: US-10-018-706-2

Sequence: 1 MVTIAINSONKRIKRLGH.....LFEFRISRGVYDPLTLVK 322

Scoring table: OLIGO

Gapop 60.0 , Gapept 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.5	164	2	A69865 conserved hypochet
2	8	2.5	227	2	H84135 hypothetical prote
3	8	2.5	244	2	S50385 hypothetical prote
4	8	2.5	306	2	T52340 cell wall-plasma m
5	8	2.5	376	2	S71558 probable cell wall
6	8	2.5	968	2	S63211 cytoskeleton assem
7	7	2.2	101	2	H86885 50S ribosomal prot
8	7	2.2	115	2	S69849 hypothetical prote
9	7	2.2	139	2	T11639 ribosomal protein
10	7	2.2	141	2	T29781 hypothetical prote
11	7	2.2	161	2	T31531 hypothetical prote
12	7	2.2	177	2	T27153 hypothetical prote
13	7	2.2	189	2	A10213 probable membrane
14	7	2.2	198	2	UC6024 kinb sporulation 8
15	7	2.2	208	2	G86352 dedd protein - Bsc
16	7	2.2	211	1	XMECD1 probable membrane
17	7	2.2	211	2	T35071 conserved hypochet
18	7	2.2	225	2	A69305 hypothetical prote
19	7	2.2	227	2	S26440 conserved hypochet
20	7	2.2	241	2	H82435 ABC transport prot
21	7	2.2	249	2	C84340 hypothetical prote
22	7	2.2	259	2	C70151 protein F55A4.4 [i
23	7	2.2	269	2	S89455 hypothetical prote
24	7	2.2	283	2	S57134 hypothetical prote
25	7	2.2	284	2	H98073 UTP-glucose-1-phos
26	7	2.2	285	2	D69184 hypothetical prote
27	7	2.2	288	2	JH0204 hydroxyproline-ric
28	7	2.2	303	2	S28264 comL, competence 1
29	7	2.2	309	2	AE3325

30	7	2.2	315	2	G84936
31	7	2.2	326	2	T26886
32	7	2.2	338	2	J00985
33	7	2.2	339	2	T40981
34	7	2.2	338	2	I40448
35	7	2.2	339	2	T34158
36	7	2.2	346	2	S19129
37	7	2.2	350	2	S22456
38	7	2.2	351	2	AC0302
39	7	2.2	355	2	G71958
40	7	2.2	367	2	H83202
41	7	2.2	368	2	UN0777
42	7	2.2	368	2	T26269
43	7	2.2	369	2	S20500
44	7	2.2	377	2	G97314
45	7	2.2	381	2	T27806
46	7	2.2	387	1	RSB74E
47	7	2.2	387	2	AD2335
48	7	2.2	387	2	T31748
49	7	2.2	391	2	B82668
50	7	2.2	391	2	S32905
51	7	2.2	409	2	T35155
52	7	2.2	413	2	AH2743
53	7	2.2	414	2	T13436
54	7	2.2	416	2	G97524
55	7	2.2	421	2	H84807
56	7	2.2	422	2	T09742
57	7	2.2	424	2	P95330
58	7	2.2	429	2	AE2501
59	7	2.2	433	2	T46528
60	7	2.2	449	2	S67819
61	7	2.2	471	2	S30585
62	7	2.2	476	2	T80182
63	7	2.2	478	1	S73920
64	7	2.2	487	2	I80183
65	7	2.2	499	2	A11107
66	7	2.2	505	2	S13859
67	7	2.2	505	2	I53417
68	7	2.2	513	2	T40998
69	7	2.2	517	2	T27927
70	7	2.2	532	2	SI8643
71	7	2.2	542	2	T29707
72	7	2.2	551	1	A55582
73	7	2.2	580	1	MMBEM5
74	7	2.2	603	2	F72237
75	7	2.2	618	1	PRECT4
76	7	2.2	618	2	H90937
77	7	2.2	618	2	D85786
78	7	2.2	623	2	S45899
79	7	2.2	623	2	T18711
80	7	2.2	634	2	T18702
81	7	2.2	635	2	T06537
82	7	2.2	652	2	E97857
83	7	2.2	729	2	B41670
84	7	2.2	731	2	T16524
85	7	2.2	772	2	T08826
86	7	2.2	781	2	T18693
87	7	2.2	814	2	T49207
88	7	2.2	836	2	T33222
89	7	2.2	865	2	A47282
90	7	2.2	873	2	A47283
91	7	2.2	886	2	S07132
92	7	2.2	887	2	AD2009
93	7	2.2	887	2	A59246
94	7	2.2	1047	2	AB2239
95	7	2.2	1119	2	C89874
96	7	2.2	1248	2	C89874
97	7	2.2	1305	2	AH3452
98	7	2.2	1336	2	T02736
99	7	2.2	1388	2	A53317
100	7	2.2	1428	2	T08852
101	7	2.2	1672	2	C81675
102	7	2.2	1965	2	T33216

kinase [imported]
hypothetical prote
hydroxyproline-ric
probable quinone o
conserved hypochet
hypothetical prote
proline-rich prote
hydroxyproline-ric
conserved hypochet
hypothetical prote
poly(beta-d-mannur
poly(beta-d-mannur
hypothetical prote
hydroxyproline-ric
alcohol dehydrogen
hypothetical prote
ribosomal protein
ornithine acetyl t
dihydrolipoamide S
serine proteinase
hypothetical prote
conserved hypochet
hypothetical prote
hypothetical prote
probable phospholi
drought-induced pr
hypothetical prote
hypothetical prote
probable GTP-4-ket
Gunc protein - Xan
hypothetical prote
activin type I rec
amidase homolog 60
activin type I rec
intercalin B [lipo
activin A receptor
type I serine-thre
hypothetical prote
hypothetical prote
maternal exupertant
probable exported
hypothetical prote
cytochrome-c oxida
U25 protein - hum
conserved hypochet
proteinase IV (EC
proteinase IV, a lmp
proteinase IV, a s
probable membrane
hypothetical prote
hypothetical prote
ethylene receptor
cell surface antig
carbon-monoxide de
hypothetical prote
hypothetical prote
receptor kinase-1i
hypothetical prote
calcium-binding pr
calpoptin - fruit
hypothetical prote
hypothetical prote
HRA protein - fru
hypothetical prote
autolysin [importe
sensory transducto
probable SCARROW
collagen alpha 1(X
lustrin A - Califo
polymorphic membra
hypothetical prote

103	7	2.2	2237	2	T21087	hypothetical prote
104	7	2.2	2245	2	T21086	hypothetical prote
105	7	2.2	2271	2	F90073	hypothetical prote
106	7	2.2	2329	2	T28115	hypothetical prote
107	7	2.2	2796	2	UC4743	fatty-acid synthas
108	7	2.2	3069	2	H70656	fatty-acid synthas
109	7	2.2	3076	2	A87058	fatty acid synthas
110	6	1.9	27	2	A43768	Hu-1like protein HB
111	6	1.9	33	2	A21211	proteasome TP14 - r
112	6	1.9	40	2	S12044	hypothetical prote
113	6	1.9	51	2	I51028	paired box transcr
114	6	1.9	56	2	T10959	proline-rich prote
115	6	1.9	57	2	S12957	venom animal kunit
116	6	1.9	57	2	C71442	hypothetical prote
117	6	1.9	60	2	AH0423	probable membrane
118	6	1.9	60	2	S10210	hypothetical prote
119	6	1.9	77	2	S47158	metallothionein II
120	6	1.9	79	2	B83400	hypothetical prote
121	6	1.9	80	2	PC4370	glutathione-disulf
122	6	1.9	83	2	G82744	hypothetical prote
123	6	1.9	84	2	G87599	hypothetical prote
124	6	1.9	84	2	T16647	hypothetical prote
125	6	1.9	85	1	MOECBH	phosphotransferase
126	6	1.9	85	1	S12749	phosphotransferase
127	6	1.9	85	1	MOEBBH	phosphotransferase
128	6	1.9	85	1	A82248	phosphocarrier pro
129	6	1.9	85	2	AB0364	PTS system, phosph
130	6	1.9	85	2	AB0364	PTS system, phosph
131	6	1.9	85	2	AB0364	PTS system, phosph
132	6	1.9	85	2	AB0364	PTS system, phosph
133	6	1.9	86	2	C41139	PTS system protein
134	6	1.9	87	1	CCFZ6	protein kinase 3 (
135	6	1.9	87	2	G90954	cytochrome c - Ap
136	6	1.9	87	2	T18152	hypothetical prote
137	6	1.9	87	2	D85803	hypothetical prote
138	6	1.9	89	2	G81106	DNA-binding protei
139	6	1.9	89	2	D82298	ribosomal protein
140	6	1.9	89	2	AB1241	ribosomal protein
141	6	1.9	89	2	AF1603	histone-like protei
142	6	1.9	89	2	UC7825	histone-like protei
143	6	1.9	95	2	A41872	heat shock protein
144	6	1.9	95	2	G97232	co-chaperonin GroB
145	6	1.9	95	2	A58889	NAMH2 dehydrogenas
146	6	1.9	98	2	PC4431	isoform 1 - bovine
147	6	1.9	99	2	T49636	related to cell di
148	6	1.9	99	2	S38868	neuropeptide HCS1
149	6	1.9	100	2	H84116	hypothetical prote
150	6	1.9	101	2	T41466	cytochrome beta
151	6	1.9	103	2	AB1032	helix-turn-helix f
152	6	1.9	103	2	B81975	probable DNA-bindi
153	6	1.9	103	2	S72925	hypothetical prote
154	6	1.9	103	2	S36545	E7 protein - human
155	6	1.9	103	2	T36213	scd protein - Str
156	6	1.9	106	2	S08315	cell wall protein
157	6	1.9	109	2	G71015	hypothetical prote
158	6	1.9	109	2	B27466	hypothetical prote
159	6	1.9	110	2	H86450	probable zinc-fing
160	6	1.9	113	2	PC4432	paired box transcr
161	6	1.9	113	2	D90171	probable integrase
162	6	1.9	114	2	B29029	pol polypeptide -
163	6	1.9	114	2	F83035	osmotically induci
164	6	1.9	115	1	JN0541	head protein gpshp
165	6	1.9	115	2	E90767	probable head deco
166	6	1.9	115	2	C90970	probable head deco
167	6	1.9	115	2	B85717	hypothetical prote
168	6	1.9	115	2	G72568	hypothetical prote
169	6	1.9	119	2	A82506	major outer membra
170	6	1.9	121	2	S05720	ribosomal protein
171	6	1.9	121	2	H72507	hypothetical prote
172	6	1.9	123	2	C75622	hypothetical prote
173	6	1.9	125	2	S78697	probable export pr
174	6	1.9	125	2	AC0753	flagellar protein
175	6	1.9	125	2	T51037	hypothetical prote
176	6	1.9	126	2	T40732	probable 50S ribos
177	6	1.9	127	2	C84502	Bn/Spm-like transp
178	6	1.9	128	2	T13112	protein gp26 - pha
179	6	1.9	129	2	S06609	surface glycoprote
180	6	1.9	130	2	S28009	root-specific prote
181	6	1.9	130	2	A18975	hypothetical prote
182	6	1.9	134	2	C87462	hypothetical prote
183	6	1.9	135	2	TS7982	T cell receptor PS
184	6	1.9	135	2	TS0933	hypothetical prote
185	6	1.9	135	2	AD1828	hypothetical prote
186	6	1.9	136	2	H83300	hypothetical prote
187	6	1.9	137	2	S35714	proline-rich prote
188	6	1.9	137	2	T22308	hypothetical prote
189	6	1.9	138	2	H87362	organic hydropox
190	6	1.9	138	2	T29558	hypothetical prote
191	6	1.9	139	2	A90454	conserved hypothet
192	6	1.9	140	2	G70664	probable thioester
193	6	1.9	141	2	T09546	extensin like prot
194	6	1.9	142	2	E86648	reductase [impor
195	6	1.9	142	2	S51582	galactoside 2-alph
196	6	1.9	144	2	AF1780	hypothetical prote
197	6	1.9	144	2	AG1404	hypothetical prote
198	6	1.9	147	2	S42552	proline-rich prote
199	6	1.9	148	2	T02305	probable membrane
200	6	1.9	148	2	G86642	hypothetical prote
201	6	1.9	148	2	F71454	protein C54B4.4 [1
202	6	1.9	148	2	D97218	hypothetical prote
203	6	1.9	149	2	D69132	probable membrane
204	6	1.9	151	2	G69345	nucleoside-diphosp
205	6	1.9	151	2	T38090	nucleoside-diphosp
206	6	1.9	151	2	G70819	hypothetical prote
207	6	1.9	154	2	T41831	hypothetical prote
208	6	1.9	154	2	G70819	hypothetical prote
209	6	1.9	155	2	T25980	hypothetical prote
210	6	1.9	155	2	E82843	hypothetical prote
211	6	1.9	156	2	RS85H2	ribosomal protein
212	6	1.9	156	2	AB4322	50S ribosomal prot
213	6	1.9	157	1	EDX73	immediate-early pr
214	6	1.9	159	2	T15173	luteinizing hormon
215	6	1.9	159	2	T48837	hypothetical prote
216	6	1.9	159	2	S65994	YybC protein - Bac
217	6	1.9	159	2	E69531	molYdoperoxin-guan
218	6	1.9	159	2	T48209	hypothetical prote
219	6	1.9	159	2	S46494	galactoside 2-alph
220	6	1.9	160	2	B71176	hypothetical prote
221	6	1.9	161	2	A83110	hypothetical prote
222	6	1.9	161	2	A95049	hypothetical prote
223	6	1.9	161	2	G97919	acetyl-CoA carboxy
224	6	1.9	161	2	H81808	hypothetical prote
225	6	1.9	164	2	T04115	isocitrate lyase (
226	6	1.9	164	2	AF3294	ompl6 protein - Br
227	6	1.9	168	2	A28146	peptidoglycan-asso
228	6	1.9	169	2	D82284	neurotensin / neur
229	6	1.9	169	2	D82284	probable acetoin u
230	6	1.9	169	2	T18321	hypothetical prote
231	6	1.9	169	2	E87512	dihydrofolate redu
232	6	1.9	171	2	G81656	conserved hypothet
233	6	1.9	174	2	H71498	hypothetical prote
234	6	1.9	174	2	E82900	hypothetical prote
235	6	1.9	175	2	G84254	hypothetical prote
236	6	1.9	176	2	AC0635	probable membrane
237	6	1.9	176	2	T82478	hypothetical prote
238	6	1.9	176	2	T43308	hypothetical prote
239	6	1.9	178	2	C75059	hypothetical prote
240	6	1.9	178	2	D72507	hypothetical prote
241	6	1.9	179	2	G75462	ribonuclease H - D
242	6	1.9	179	2	A64846	probable membrane
243	6	1.9	179	2	E90805	hypothetical prote
244	6	1.9	179	2	B85665	hypothetical prote
245	6	1.9	181	2	T44972	gas-vesicle operon
246	6	1.9	184	2	AE0071	probable lipoprote
247	6	1.9	185	2	S60778	dhhydridipicolinat
248	6	1.9	185	2	H81837	peptidylprolyl iso

249	6	1.9	185	2	H81103	peptidylprolyl iso	322	6	1.9	232	2	B81339	probable flagellar
250	6	1.9	185	2	C70550	hypothetical prote	323	6	1.9	233	1	C39142	mkx protein precu
251	6	1.9	186	2	C72782	hypothetical prote	324	6	1.9	233	2	T32500	hypothetical prote
252	6	1.9	187	1	B29830	toluene dioxygenas	325	6	1.9	234	2	AG1320	glycerol uptake fa
253	6	1.9	187	2	S76728	hypothetical prote	326	6	1.9	234	2	S60885	terric exochelin u
254	6	1.9	187	2	T50946	hypothetical prote	327	6	1.9	234	2	S65073	protein disulfide
255	6	1.9	188	2	S74560	hypothetical prote	328	6	1.9	236	2	E91099	protein disulfide
256	6	1.9	188	2	D87705	intracellular sept	329	6	1.9	236	2	A85945	hypothetical prote
257	6	1.9	190	2	S75069	K+-transporting AT	330	6	1.9	236	2	T40239	hypothetical prote
258	6	1.9	191	2	T04953	hypothetical prote	331	6	1.9	237	2	E84311	hypothetical prote
259	6	1.9	191	2	AH0196	probable lipoprote	332	6	1.9	239	2	D70854	hypothetical prote
260	6	1.9	192	2	F70005	hypothetical prote	333	6	1.9	241	2	B49202	cysteine-rich secr
261	6	1.9	192	2	A10383	probable lipoprote	334	6	1.9	241	2	T05040	nodulin-26-like pr
262	6	1.9	192	2	E96940	uncharacterized me	335	6	1.9	241	2	H72691	probable hexulose
263	6	1.9	194	2	J01095	hypothetical 21.8x	336	6	1.9	242	2	T45421	probable 2-hydroxy
264	6	1.9	194	2	E90223	rRNA adenine N-6-m	337	6	1.9	242	2	A82754	lipoprotein Xf0855
265	6	1.9	194	2	JC7252	flavoredoxin FR I	338	6	1.9	243	2	S44914	Zk68.7 protein -
266	6	1.9	195	2	E64755	YagZ protein - Esc	339	6	1.9	244	2	S73022	probable membrane
267	6	1.9	195	2	C90669	hypothetical prote	340	6	1.9	244	2	C97532	transcription regu
268	6	1.9	195	2	F85519	hypothetical prote	341	6	1.9	244	2	AC2751	hypothetical prote
269	6	1.9	195	2	A13398	probable capsule p	342	6	1.9	246	2	C84998	hypothetical prote
270	6	1.9	196	2	AC3300	conserved membrane-	343	6	1.9	248	2	T26461	hypothetical prote
271	6	1.9	196	2	T64222	conserved hypochet	344	6	1.9	248	2	T33230	hypothetical prote
272	6	1.9	197	2	C72545	probable 3-octapre	345	6	1.9	248	2	T02122	hypothetical prote
273	6	1.9	197	2	T36071	probable pantoate-	346	6	1.9	249	2	T25643	hypothetical prote
274	6	1.9	197	2	S53053	hypothetical prote	347	6	1.9	249	2	JQ0133	hypothetical prote
275	6	1.9	198	2	D72603	hypothetical prote	348	6	1.9	250	2	A28564	hypothetical 26.4k
276	6	1.9	199	2	E69111	conserved hypochet	349	6	1.9	251	2	D71931	lymphocyte functio
277	6	1.9	200	1	Q08E27	BALF1 protein - hu	350	6	1.9	252	2	H75123	hypothetical prote
278	6	1.9	200	1	D64008	hypothetical prote	351	6	1.9	253	2	S75731	thiamin biosynthe
279	6	1.9	203	2	JC2207	LamI protein - tr	352	6	1.9	256	2	A84862	3-isopropylmalate
280	6	1.9	203	2	H83963	guanylate kinase g	353	6	1.9	257	2	JU0403	hypothetical 28.1k
281	6	1.9	205	1	A65094	YglH protein - Esc	354	6	1.9	257	2	A10844	glucitol operon re
282	6	1.9	205	2	F91121	hypothetical prote	355	6	1.9	257	2	S76930	hypothetical prote
283	6	1.9	205	2	E85966	hypothetical prote	356	6	1.9	257	2	A82680	conserved hypochet
284	6	1.9	205	2	B70168	conserved hypochet	357	6	1.9	258	2	C86286	protein Pg1.16 [I
285	6	1.9	205	2	D75178	hypothetical prote	358	6	1.9	258	2	S23106	p1A protein - shi
286	6	1.9	207	2	G75118	proteasome, chain	359	6	1.9	258	2	A81265	probable UDP-N-ace
287	6	1.9	207	2	A83540	phospholipase acce	360	6	1.9	259	2	G87083	probable rRNA meth
288	6	1.9	207	2	AG0055	probable membrane	361	6	1.9	259	2	E87537	hypothetical prote
289	6	1.9	209	2	H85644	hypothetical prote	362	6	1.9	259	2	G83057	hypothetical prote
290	6	1.9	209	2	C90785	probable anti-repr	363	6	1.9	260	2	G70619	probable 23s rnam
291	6	1.9	211	2	G82573	CDP-diacylglycerol	364	6	1.9	260	2	T26174	hypothetical prote
292	6	1.9	211	2	S21864	probable cysteine	365	6	1.9	260	2	A99290	hypothetical prote
293	6	1.9	214	2	G83592	hypothetical prote	366	6	1.9	261	2	B64746	YafT protein - Esc
294	6	1.9	215	2	SS4293	keratin 6 - bovine	367	6	1.9	261	2	C97462	ATP synthase chain
295	6	1.9	215	2	C72759	hypothetical prote	368	6	1.9	262	2	S41753	acyl-lactyl-carrier
296	6	1.9	217	2	S10212	late 33K protein -	369	6	1.9	262	2	C97426	flagellar basal-bo
297	6	1.9	218	2	E82780	endonuclease III x	370	6	1.9	262	2	AC2644	flagellar basal-bo
298	6	1.9	218	2	AG3486	hemolysin III (imp	371	6	1.9	262	2	S60213	fomc protein - Str
299	6	1.9	219	1	HSX118	histone H1B - Afri	372	6	1.9	262	2	AD2819	conserved hypochet
300	6	1.9	219	2	A75628	response regulator	373	6	1.9	264	2	D83522	probable radical a
301	6	1.9	220	2	T51447	histone H1B - Afri	374	6	1.9	265	2	A43899	transcription fact
302	6	1.9	221	2	G90982	acetyltransferase	375	6	1.9	265	2	S77181	hypothetical prote
303	6	1.9	221	2	E85828	acetyl transferase	376	6	1.9	266	2	T35800	probable polyamine
304	6	1.9	222	2	E84283	TRK potassium upla	377	6	1.9	266	2	E69083	thiamin biosynthes
305	6	1.9	223	2	B97568	trp repressor bind	378	6	1.9	266	2	T07363	hypothetical prote
306	6	1.9	223	2	A12788	flavodoxin (import	379	6	1.9	266	2	D85506	hypothetical prote
307	6	1.9	223	2	C71287	probable hypocheti	380	6	1.9	266	2	T37878	hypothetical prote
308	6	1.9	224	2	D72861	AcOxI-91 protein -	381	6	1.9	267	2	S08314	cell wall glycopro
309	6	1.9	224	2	T21788	hypothetical prote	382	6	1.9	267	2	F90028	hypothetical prote
310	6	1.9	226	2	T05976	hypothetical prote	383	6	1.9	268	2	T05819	hypothetical prote
311	6	1.9	226	2	T29210	hypothetical prote	384	6	1.9	268	2	A81424	cytolethal distend
312	6	1.9	226	2	B95362	protein (imported	385	6	1.9	269	2	A99396	conserved hypochet
313	6	1.9	228	2	G83115	30S ribosomal prot	386	6	1.9	270	2	T47853	hypothetical prote
314	6	1.9	228	2	G69405	conserved hypochet	387	6	1.9	272	2	T41070	conserved hypochet
315	6	1.9	229	2	A97349	uncharacterized co	388	6	1.9	273	2	AH0278	probable transpor
316	6	1.9	229	2	F70568	hypothetical prote	389	6	1.9	273	2	AD2322	ABC nitrate transp
317	6	1.9	231	2	C86665	amino acid ABC tra	390	6	1.9	273	2	S64054	hypothetical prote
318	6	1.9	231	2	C70595	hypothetical prote	391	6	1.9	273	2	S67622	hypothetical prote
319	6	1.9	231	2	C64703	hypothetical prote	392	6	1.9	274	2	T47856	hypothetical prote
320	6	1.9	231	2	C71816	probable outer mem	393	6	1.9	275	2	T43004	hypothetical prote
321	6	1.9	232	2	A10019	phosphoglycolate p	394	6	1.9	276	2	S23891	secretory pathway

395	6	1.9	276	2	H75169	468	6	1.9	305	2	AG1014	malose operon per
396	6	1.9	276	2	B87203	469	6	1.9	305	2	D75551	probable endo alpha
397	6	1.9	276	2	B75007	470	6	1.9	306	2	D83992	proline oxidase (p
398	6	1.9	277	2	T21630	471	6	1.9	306	2	AB3049	arginase [imported
399	6	1.9	277	2	AB1883	472	6	1.9	307	2	D87688	hypothetical prote
400	6	1.9	277	2	H87553	473	6	1.9	307	2	S37252	homeotic protein H
401	6	1.9	278	2	S57351	474	6	1.9	307	2	P84502	Bn/Gpm transposon
402	6	1.9	278	2	A75427	475	6	1.9	308	2	H85214	nodulin-26-like pr
403	6	1.9	279	2	D90080	476	6	1.9	308	2	G71110	hypothetical prote
404	6	1.9	280	2	T02471	477	6	1.9	308	2	B82789	hypothetical prote
405	6	1.9	280	2	E96643	478	6	1.9	309	2	E97175	hypothetical prote
406	6	1.9	281	2	D97100	479	6	1.9	309	2	G90500	cysteine synthase
407	6	1.9	281	2	C84868	480	6	1.9	310	2	P84098	Hpr (Ser/Thr) prot
408	6	1.9	281	2	A10461	481	6	1.9	310	2	T20535	hypothetical prote
409	6	1.9	283	2	B87603	482	6	1.9	311	2	E72277	phosphotibodyform
410	6	1.9	283	2	S13383	483	6	1.9	312	2	AE1385	Hpr-P (Ser) kinase/
411	6	1.9	284	2	S25140	484	6	1.9	312	2	T36558	hypothetical prote
412	6	1.9	284	2	B41224	485	6	1.9	312	2	T36558	hypothetical prote
413	6	1.9	284	2	B90190	486	6	1.9	312	2	T16001	RNA-directed RNA p
414	6	1.9	284	2	G86477	487	6	1.9	312	4	A40473	probable flavoprot
415	6	1.9	284	2	AE0314	488	6	1.9	313	2	E90634	probable flavoprot
416	6	1.9	286	2	T47064	489	6	1.9	313	2	E85485	electron transfer
417	6	1.9	286	2	A10230	489	6	1.9	313	2	AF0511	hypothetical prote
418	6	1.9	287	2	T48548	490	6	1.9	313	2	B64725	hypothetical prote
419	6	1.9	288	2	A13217	491	6	1.9	313	2	T39581	hypothetical prote
420	6	1.9	288	2	D84239	492	6	1.9	313	2	S33027	hypothetical prote
421	6	1.9	288	2	C83487	493	6	1.9	314	2	E84046	hypothetical prote
422	6	1.9	288	2	S40173	494	6	1.9	314	2	S75685	hypothetical prote
423	6	1.9	291	2	S48977	495	6	1.9	314	2	T39756	hypothetical prote
424	6	1.9	291	2	F82897	496	6	1.9	314	2	T28879	ADP-hepse synth
425	6	1.9	292	2	AC1460	497	6	1.9	315	2	H70350	ADP-hepse synth
426	6	1.9	292	2	AD1097	498	6	1.9	315	2	D91254	regulatory protein
427	6	1.9	292	2	HE1449	499	6	1.9	315	2	S50185	hypothetical prote
428	6	1.9	293	2	BE1468	500	6	1.9	316	2	T31880	protein B0238.11 l
429	6	1.9	293	2	B80865	501	6	1.9	317	2	G89044	cysteine synthase
430	6	1.9	293	2	G85753	502	6	1.9	320	2	AH2374	GDP-D-mannose dehy
431	6	1.9	293	2	BE4880	503	6	1.9	320	2	S74773	comE-like protein
432	6	1.9	293	2	F70724	504	6	1.9	320	2	T37990	probable polysacch
433	6	1.9	293	2	B87632	505	6	1.9	321	2	G43691	catechol 2,3-dioxy
434	6	1.9	293	2	G98163	506	6	1.9	321	2	DE9101	CesTymA F420-dep
435	6	1.9	293	2	AB3124	507	6	1.9	321	2	DE9101	NS,N10-methylamete
436	6	1.9	293	2	C91105	508	6	1.9	321	2	B65529	probable 60S ribos
437	6	1.9	294	2	F85950	509	6	1.9	321	2	AE1611	UDP-N-acetylglucosyl
438	6	1.9	294	2	G65078	510	6	1.9	321	2	A97611	UDP-N-acetylglucosyl
439	6	1.9	294	2	A12016	511	6	1.9	321	2	AE2833	hypothetical prote
440	6	1.9	295	2	I40810	512	6	1.9	322	2	H86094	pyruvate dehydroge
441	6	1.9	295	2	B91166	513	6	1.9	322	2	S14230	pyruvate dehydroge
442	6	1.9	295	2	B86012	514	6	1.9	325	2	F83981	pyruvate dehydroge
443	6	1.9	295	2	S48102	515	6	1.9	325	2	AE1563	pyruvate dehydroge
444	6	1.9	295	2	AE0131	516	6	1.9	325	2	AE1506	pyruvate dehydroge
445	6	1.9	295	2	B97597	517	6	1.9	325	2	AD1708	phosphotransacetyl
446	6	1.9	295	2	F75209	518	6	1.9	325	2	AG1337	phosphotransacetyl
447	6	1.9	296	1	I39938	519	6	1.9	325	2	T17473	hypothetical prote
448	6	1.9	297	1	F65002	520	6	1.9	325	2	E35270	6-phosphofructokin
449	6	1.9	297	2	D91027	521	6	1.9	327	1	B35370	hypothetical prote
450	6	1.9	297	2	B85871	522	6	1.9	327	2	A83747	acetoaldehyde dehydrogen
451	6	1.9	298	2	D87444	523	6	1.9	327	2	A98237	arginase (U39262)
452	6	1.9	298	2	T25655	524	6	1.9	327	2	A57626	peroxisomal enoyl
453	6	1.9	298	2	T23250	525	6	1.9	327	2	D82090	conserved hypochet
454	6	1.9	299	2	AD0507	526	6	1.9	327	2	T32583	hypothetical prote
455	6	1.9	299	2	AC0507	527	6	1.9	328	2	T21985	hypothetical prote
456	6	1.9	299	2	F84785	528	6	1.9	328	2	F89827	phosphotransacetyl
457	6	1.9	299	2	T33830	529	6	1.9	328	2	B75320	hypothetical prote
458	6	1.9	299	2	A70713	530	6	1.9	328	2	S01075	chitinase (BC 3.2.
459	6	1.9	300	2	D83158	531	6	1.9	328	2	G84127	l-arabinose transp
460	6	1.9	301	1	OE0C3R	532	6	1.9	330	2	SS3717	phosphotransacetyl
461	6	1.9	301	1	B90631	533	6	1.9	330	2	T05717	hypophyrynosogen d
462	6	1.9	302	1	B85482	534	6	1.9	330	2	SS3717	probable extensin
463	6	1.9	302	1	F77021	535	6	1.9	330	2	A40855	homeotic protein H
464	6	1.9	302	1	A13645	536	6	1.9	330	2	H90184	hypothetical prote
465	6	1.9	303	2	H83849	537	6	1.9	332	1	F89771	lipoprotein (lipor
466	6	1.9	304	2	AG3324	538	6	1.9	332	1	B39136	flagellar motor sw
467	6	1.9	305	2	S20600	539	6	1.9	332	2	F64096	gal repressor - Ha
						540	6	1.9	332	2	I57032	gene Tlx-1 protein

541	6	1.9	333	2	H89961	hypothetical prote	614	6	1.9	370	2	E82751	lipopolysaccharide
542	6	1.9	334	2	T43865	phosphate transpor	615	6	1.9	370	2	AB3334	daunorubicin resis
543	6	1.9	T22215	2	hypothetical prote	616	6	1.9	372	2	S76427	hypothetical prote	
544	6	1.9	335	2	G82936	prolipoprotein dia	617	6	1.9	372	2	A98157	probable permease
545	6	1.9	335	2	T03229	probable chitinase	618	6	1.9	372	2	AH3130	ABC transporter, m
546	6	1.9	335	2	S03212	hypothetical prote	619	6	1.9	373	2	B82697	rod shape-determ
547	6	1.9	336	2	S70255	flagellin - Lyme d	620	6	1.9	373	2	A56392	beta-galactoside a
548	6	1.9	336	2	I40204	flagellin - Borrel	621	6	1.9	374	2	S21736	translocating chai
549	6	1.9	B72617	2	probable dipeptid	622	6	1.9	374	2	C91198	Espp protein [limo	
550	6	1.9	335	2	B49683	prsh 3'-region pro	623	6	1.9	374	2	G86044	secreted protein E
551	6	1.9	336	2	A87300	TPR domain protein	624	6	1.9	374	2	C88734	protein F32B10.6 l
552	6	1.9	336	2	H75479	lipocate-protein 11	625	6	1.9	374	2	AC0983	hypothetical prote
553	6	1.9	336	2	T22767	hypothetical prote	626	6	1.9	375	2	A32827	retun precursor -
554	6	1.9	337	2	G95915	probable sugar ABC	627	6	1.9	375	2	A64187	xylose transport p
555	6	1.9	337	2	S37799	hypothetical prote	628	6	1.9	376	2	S04497	surface antigen PA
556	6	1.9	338	2	D90697	adhesin/invasin-11	629	6	1.9	376	2	C90444	hypothetical prote
557	6	1.9	338	2	G85547	adhesin/invasin-11	630	6	1.9	376	2	H82988	hypothetical prote
558	6	1.9	339	2	A50754	hypothetical prote	631	6	1.9	377	1	B0AG58	virB10 protein - A
559	6	1.9	340	2	S40414	chitinase (EC 3.2.	632	6	1.9	377	2	AF3249	component of type
560	6	1.9	340	2	T45805	glucan endo-1,3-be	633	6	1.9	379	2	C75006	hypothetical prote
561	6	1.9	341	2	H90797	probable tellurium	634	6	1.9	381	2	S24511	latent membrane pr
562	6	1.9	342	2	S63356	probable membrane	635	6	1.9	381	2	S52985	cell wall protein
563	6	1.9	342	2	D49348	succinoglycan bios	636	6	1.9	381	2	F87553	aminoctransferase,
564	6	1.9	342	2	B64491	NS,M10-methylenete	637	6	1.9	382	2	T25538	hypothetical prote
565	6	1.9	342	2	G95975	glucosyltransferas	638	6	1.9	383	2	TG5BEKG	glycoprotein precu
566	6	1.9	342	2	AD0346	probable transport	639	6	1.9	383	2	T35773	succinyl-Coa synth
567	6	1.9	342	2	AE2490	Wp-repeat protein	640	6	1.9	384	2	S70989	recf protein - Myc
568	6	1.9	343	2	A56098	alpha(1,2)fucosylt	641	6	1.9	384	2	S74774	hypothetical prote
569	6	1.9	344	2	G64122	phosphoribosylform	642	6	1.9	385	2	E70154	DNA-directed DNA p
570	6	1.9	344	2	S45912	Yko2 protein - Yea	643	6	1.9	385	2	S68780	dopamine D1-like r
571	6	1.9	345	1	VABPA7	major capsid prote	644	6	1.9	385	2	C83174	hypothetical prote
572	6	1.9	345	2	B86202	hypothetical prote	645	6	1.9	385	2	A85607	hypothetical prote
573	6	1.9	346	2	F81351	probable D-alanine	646	6	1.9	386	1	LABERJ	latent membrane pr
574	6	1.9	347	2	D96590	hypothetical prote	647	6	1.9	386	2	A82284	conserved hypochet
575	6	1.9	347	2	H75253	hypothetical prote	648	6	1.9	386	2	F83994	xylose operon tran
576	6	1.9	347	2	T50990	hypothetical prote	649	6	1.9	388	2	S25298	extensin (clone To
577	6	1.9	348	2	B75312	branched-chain ami	650	6	1.9	388	2	S77534	high-affinity bran
578	6	1.9	348	2	D88088	protein B0454.1 [i	651	6	1.9	389	2	H95003	aromatic amino aci
579	6	1.9	351	2	TJ00056	flagellin - Serrat	652	6	1.9	389	2	C97876	hypothetical prote
580	6	1.9	351	2	T32717	hypothetical prote	653	6	1.9	389	2	D83139	hypothetical prote
581	6	1.9	354	2	T19856	hypothetical prote	654	6	1.9	390	1	Q0BE77	glycoprotein I pre
582	6	1.9	354	2	T24094	hypothetical prote	655	6	1.9	390	2	AG2549	hypothetical prote
583	6	1.9	354	2	B56392	beta-galactoside a	656	6	1.9	390	2	G88022	protein T27A1.4 [i
584	6	1.9	354	2	AG1324	glucosamine-fructo	657	6	1.9	390	2	E84066	hypothetical prote
585	6	1.9	354	2	A11695	weakly glucosamine	658	6	1.9	391	1	S02192	cellular tumor ant
586	6	1.9	355	2	AD0299	conserved hypochet	659	6	1.9	391	2	S55732	utroporphyrinogen d
587	6	1.9	355	2	A47039	nylon oligomer-deg	660	6	1.9	391	2	B83203	arginate o-acetyl
588	6	1.9	355	2	D95270	membrane protein h	661	6	1.9	391	2	G83346	conserved hypochet
589	6	1.9	355	2	AE1642	class I histocompa	662	6	1.9	392	2	T20026	hypothetical prote
590	6	1.9	356	2	JH0289	1-allyl-threonine a	663	6	1.9	392	2	A13375	rare lipoprotein A
591	6	1.9	356	2	B89919	hypothetical prote	664	6	1.9	392	2	T44146	probable uroporphy
592	6	1.9	357	2	D84395	1-allyl-threonine a	665	6	1.9	393	2	T01653	utroporphyrinogen d
593	6	1.9	357	2	S37925	hypothetical prote	666	6	1.9	393	2	T45434	probable succinyl-
594	6	1.9	359	2	S55032	dpy-20 protein - C	667	6	1.9	393	2	S69537	hypothetical prote
595	6	1.9	359	2	B72290	branched chain ami	668	6	1.9	393	2	E97248	PIP-dependent amin
596	6	1.9	359	2	S42787	serine/threonine-r	669	6	1.9	393	2	B97447	hypothetical prote
597	6	1.9	362	2	JH0291	class I histocompa	670	6	1.9	394	2	B84830	probable uroporphy
598	6	1.9	362	2	JH0292	class I histocompa	671	6	1.9	395	2	B75512	conserved hypochet
599	6	1.9	362	2	A12878	hypothetical prote	672	6	1.9	395	2	A54949	syndecan precursor
600	6	1.9	362	2	C97655	probable hemin ABC	673	6	1.9	395	2	S73732	MC306 homolog A05-
601	6	1.9	362	2	T33092	hypothetical prote	674	6	1.9	395	2	F72424	hypothetical prote
602	6	1.9	363	2	S76156	hypothetical prote	675	6	1.9	396	2	A91019	probable antibiotic
603	6	1.9	363	2	F84360	threonine synthase	676	6	1.9	396	2	C85863	probable antibioci
604	6	1.9	364	1	JE0292	fungai stress prot	677	6	1.9	396	2	AF0791	probable transmemb
605	6	1.9	365	2	T49733	probable homoserin	678	6	1.9	397	2	G95850	probable efflux pr
606	6	1.9	365	2	A43720	parathion hydrolas	679	6	1.9	397	2	C82992	hypothetical prote
607	6	1.9	365	2	B75398	hypothetical prote	680	6	1.9	397	2	T44437	hypothetical prote
608	6	1.9	365	2	A36047	galactoside 2-alph	681	6	1.9	397	2	T40838	probable transpor
609	6	1.9	366	1	W2MLR1	E2 protein - rhesu	682	6	1.9	398	1	VBBPA7	minor capsid prote
610	6	1.9	367	2	A13319	choleyglycine hyd	683	6	1.9	398	2	T20669	hypothetical prote
611	6	1.9	368	2	T24983	hypothetical prote	684	6	1.9	400	2	H84492	En/Spm transposon
612	6	1.9	368	2	G88636	protein W09G12.7 [685	6	1.9	402	2	JE0282	cell division prot
613	6	1.9	369	2	AC2665	conserved hypochet	686	6	1.9	402	2	AD3451	hypothetical prote

687	6	1.9	403	2	H86976	760	6	1.9	438	2	E55578	hypothetical prote
688	6	1.9	403	2	H71484	761	6	1.9	439	2	G75314	preproteic translo
689	6	1.9	404	1	LA55CA	762	6	1.9	439	2	G70693	probable dnf prot
690	6	1.9	404	2	LA5577	763	6	1.9	439	2	E82983	probable oxidoredu
691	6	1.9	405	2	H82010	764	6	1.9	439	2	S51939	chitinase (EC 3.2.
692	6	1.9	405	2	G91020	765	6	1.9	440	2	G87444	8-amin-7-oxonon
693	6	1.9	405	2	F85864	766	6	1.9	440	2	A56693	receptor protein k
694	6	1.9	405	2	AB0793	767	6	1.9	441	2	G64094	preproteic translo
695	6	1.9	405	2	E64995	768	6	1.9	441	2	T12011	cellulase (EC 3.2.
696	6	1.9	405	4	A61181	769	6	1.9	442	2	S50062	cell wall glycopro
697	6	1.9	407	2	S25841	770	6	1.9	443	2	D97618	probable mfs trans
698	6	1.9	407	2	T28795	771	6	1.9	443	2	AB2841	MFS permease/proi
699	6	1.9	407	2	H87341	772	6	1.9	443	2	DB1233	hypothetical prote
700	6	1.9	407	2	D95882	773	6	1.9	444	2	T45602	cellobiose/cellotr
701	6	1.9	408	2	A86652	774	6	1.9	444	2	AS5871	hydrogen peroxide-
702	6	1.9	409	1	B0AC55	775	6	1.9	445	2	T30604	hypothetical prote
703	6	1.9	409	1	I48607	776	6	1.9	445	2	B83931	hypothetical prote
704	6	1.9	410	2	T23264	777	6	1.9	447	1	T43319	hypothetical prote
705	6	1.9	411	2	S41648	778	6	1.9	447	2	T05070	hypothetical prote
706	6	1.9	411	2	E85919	779	6	1.9	447	2	B83919	hypothetical prote
707	6	1.9	411	2	F91074	780	6	1.9	448	2	S44133	hypothetical prote
708	6	1.9	412	2	H95313	781	6	1.9	448	2	DB5362	probable oxidoredu
709	6	1.9	412	2	E97103	782	6	1.9	448	2	G83333	probable porin PA2
710	6	1.9	414	2	E84266	783	6	1.9	449	2	S12382	glycine receptor a
711	6	1.9	414	2	C75461	784	6	1.9	449	2	E70016	purine permease ho
712	6	1.9	414	2	T33067	785	6	1.9	449	2	T33257	hypothetical prote
713	6	1.9	415	2	E89467	786	6	1.9	450	2	F82848	gtp-binding protei
714	6	1.9	415	2	H81079	787	6	1.9	451	2	C49970	glycine receptor a
715	6	1.9	416	2	G81864	788	6	1.9	451	2	AB2296	UDP-N-acetylglucos
716	6	1.9	416	2	T45051	789	6	1.9	451	2	C49898	cellobiose phospho
717	6	1.9	416	2	H71493	790	6	1.9	451	2	T41877	MES3 orf139 - Bomb
718	6	1.9	416	2	B84710	791	6	1.9	451	2	AF3453	probable glucosylt
719	6	1.9	417	2	T30963	792	6	1.9	452	2	T00981	protein T13P2.1 [l
720	6	1.9	418	2	S11678	793	6	1.9	454	2	H88791	probable cell wall
721	6	1.9	419	2	A70156	794	6	1.9	454	2	E75291	UDP-N-acetylglucos
722	6	1.9	419	2	H71662	795	6	1.9	456	2	S66080	NADH dehydrogenas
723	6	1.9	420	2	S50652	796	6	1.9	457	2	A53669	streptogristin C (E
724	6	1.9	420	2	A12458	797	6	1.9	457	2	S20662	glycine receptor a
725	6	1.9	421	2	G97698	798	6	1.9	457	2	H83662	DNA repair protein
726	6	1.9	421	2	AH2924	799	6	1.9	458	1	S74543	transcription term
727	6	1.9	421	2	AF3252	800	6	1.9	458	1	T33777	glutathione-disulf
728	6	1.9	421	2	AF3252	801	6	1.9	458	2	B84048	argininosuccinate
729	6	1.9	421	2	G98158	802	6	1.9	458	2	AB2426	glutathione reduct
730	6	1.9	421	2	AB3129	803	6	1.9	459	2	AH2426	regulatory protein
731	6	1.9	422	1	A56674	804	6	1.9	460	2	G90035	tear protein limpo
732	6	1.9	423	2	C81266	805	6	1.9	460	2	AC2274	glycosyltransferas
733	6	1.9	423	2	H83484	806	6	1.9	462	2	G82964	cytochrome P450 [l
734	6	1.9	426	2	D88103	807	6	1.9	462	2	H69751	amino acid permeas
735	6	1.9	427	1	S63615	808	6	1.9	463	1	GRECNK	nitrite extrusion
736	6	1.9	427	2	S66046	809	6	1.9	463	2	H90844	nitrite extrusion
737	6	1.9	428	2	JH0634	810	6	1.9	463	2	G85702	nitrite extrusion
738	6	1.9	428	2	B69187	811	6	1.9	464	2	T39699	glutathione-disulf
739	6	1.9	428	2	T32952	812	6	1.9	464	2	A47023	S-layer protein -
740	6	1.9	428	2	C97075	813	6	1.9	465	2	S39494	glutathione-disulf
741	6	1.9	428	2	AB3550	814	6	1.9	466	2	T35164	probable secreted
742	6	1.9	429	2	AB3550	815	6	1.9	466	2	A12231	glutamate-ammonia
743	6	1.9	430	2	G70581	816	6	1.9	467	2	S33181	inositol 1,4,5-tri
744	6	1.9	431	2	T29267	817	6	1.9	468	2	A49131	probable replicati
745	6	1.9	431	2	E89357	818	6	1.9	468	2	E81876	replicative DNA he
746	6	1.9	432	1	E89357	819	6	1.9	468	2	E81876	terminase large ch
747	6	1.9	432	2	T10894	820	6	1.9	468	2	T12725	hypothetical prote
748	6	1.9	433	2	E70968	821	6	1.9	468	2	AB2294	hypothetical prote
749	6	1.9	433	2	S42234	822	6	1.9	470	2	A12188	lysostaphin (EC 3.
750	6	1.9	436	1	B70473	823	6	1.9	472	2	S18452	variant surface gl
751	6	1.9	436	2	S56529	824	6	1.9	474	2	R75276	sensor histidine k
752	6	1.9	437	2	AF0667	825	6	1.9	474	2	E69461	4-hydroxybutyrate
753	6	1.9	437	2	A34800	826	6	1.9	474	2	T50258	hypothetical coile
754	6	1.9	437	2	T39592	827	6	1.9	474	2	T39466	lipase - pseudomon
755	6	1.9	437	2	E91240	828	6	1.9	476	2	AB3942	transcription fact
756	6	1.9	438	2	B66088	829	6	1.9	477	2	E46336	hydrolase (aux2) h
757	6	1.9	438	2	D70528	830	6	1.9	477	2	E46336	
758	6	1.9	438	2	D70528	831	6	1.9	477	2	E46336	
759	6	1.9	438	2	D70528	832	6	1.9	477	2	E46336	

833	6	1.9	477	2	T46304	hypothetical prote
834	6	1.9	479	1	RDHU	glutathione-disulf
835	6	1.9	479	1	B65051	hypothetical prote
836	6	1.9	479	2	EG7337	aspartyl/asparagin
837	6	1.9	479	2	AD0845	probable flavoprot
838	6	1.9	479	2	A64117	serine-type D-Ala-
839	6	1.9	480	2	AC2440	hypothetical prote
840	6	1.9	480	2	H71430	hypothetical prote
841	6	1.9	480	2	JC7506	heparanase protein
842	6	1.9	480	2	A35282	alpha-amylase (EC
843	6	1.9	484	2	JE0261	N-acetylglucosamin
844	6	1.9	484	2	B82110	conserved hypochet
845	6	1.9	485	2	T32538	hypothetical prote
846	6	1.9	485	2	T28076	hypothetical prote
847	6	1.9	486	2	S66097	cell-cycle protein
848	6	1.9	488	2	P84493	hypothetical prote
849	6	1.9	489	2	G64039	hypothetical prote
850	6	1.9	490	2	T41646	hypothetical prote
851	6	1.9	490	2	T39534	hypothetical prote
852	6	1.9	491	2	C46230	RNA-binding protei
853	6	1.9	491	2	S61306	LEFT protein homol
854	6	1.9	493	2	AD1398	drug-export protei
855	6	1.9	493	2	AG1773	drug-export protei
856	6	1.9	494	2	F72095	na-dependent trans
857	6	1.9	494	2	B86527	na-dependent trans
858	6	1.9	496	2	D75261	conserved hypochet
859	6	1.9	496	2	T30976	hypothetical prote
860	6	1.9	498	1	VH1YN6	nucleoside-protein
861	6	1.9	498	1	VH1V43	nucleoside-protein
862	6	1.9	498	1	A96541	probable hexokinase
863	6	1.9	499	1	S66677	chirodoxin-disulf
864	6	1.9	499	1	S26454	hypothetical prote
865	6	1.9	499	2	JC2062	transforming growt
866	6	1.9	500	1	S60829	probable aldehyde
867	6	1.9	500	2	A48053	phosphatidylserine
868	6	1.9	501	2	S48120	deoxyribodipyrimid
869	6	1.9	502	2	T35910	probable carboxyle
870	6	1.9	502	2	H82832	glucose-6-phosphat
871	6	1.9	502	2	T02746	cyclin A-like prot
872	6	1.9	502	2	A10151	probable bacteriop
873	6	1.9	503	2	A49432	activin receptor-1
874	6	1.9	503	2	JC2061	transforming growt
875	6	1.9	504	2	F70813	hypothetical prote
876	6	1.9	504	2	A81922	probable cell-surf
877	6	1.9	504	2	F72744	hypothetical prote
878	6	1.9	505	2	T28276	ORF MSV15 probabl
879	6	1.9	507	2	T31179	hypothetical prote
880	6	1.9	508	2	C87564	cell division proc
881	6	1.9	508	2	D81325	probable phosphate
882	6	1.9	509	2	A45992	activin A receptor
883	6	1.9	509	2	T59576	transforming growt
884	6	1.9	509	2	A49664	activin type I rec
885	6	1.9	509	2	T32201	hypothetical prote
886	6	1.9	510	1	S15620	L1 protein - human
887	6	1.9	510	1	S15627	L1 protein - human
888	6	1.9	510	2	T39601	serine carboxypept
889	6	1.9	510	2	A45338	connexin-56 - chic
890	6	1.9	511	2	S44275	dopamine receptor
891	6	1.9	511	2	JE0312	3-ketosteroid-deta
892	6	1.9	512	2	AD0416	probable type II s
893	6	1.9	514	2	A87266	senor histidine k
894	6	1.9	515	2	AH2396	glycerol-3-phospha
895	6	1.9	516	2	T50190	probable phosphati
896	6	1.9	518	1	P2WL8	L2 protein - human
897	6	1.9	518	1	P2WL5	L2 protein - human
898	6	1.9	518	1	P2WL5	L2 protein - human
899	6	1.9	518	1	P2WL5	L2 protein - human
900	6	1.9	518	1	P2WL5	L2 protein - human
901	6	1.9	519	2	S36471	L2 protein - human
902	6	1.9	519	2	S36471	L2 protein - human
903	6	1.9	520	2	S36489	GAGA transcription
904	6	1.9	520	2	S36495	L2 protein - human
905	6	1.9	521	2	B95899	probable ABC trans
906	6	1.9	521	2	S36571	L2 protein - human
907	6	1.9	521	2	A43319	chromaffin granule
908	6	1.9	521	2	T45608	hypothetical prote
909	6	1.9	523	1	VHNZMV	nucleocapsid prote
910	6	1.9	523	2	T30091	nucleocapsid prote
911	6	1.9	523	2	F71302	asparagine-tRNA 11
912	6	1.9	523	2	A38101	hypothetical prote
913	6	1.9	523	2	C84753	hypothetical prote
914	6	1.9	524	2	S46007	hypothetical prote
915	6	1.9	524	2	T06021	hypothetical prote
916	6	1.9	525	1	A48556	nucleocapsid prote
917	6	1.9	525	1	VHNZMV	nucleocapsid prote
918	6	1.9	525	1	JU0272	nucleocapsid prote
919	6	1.9	525	2	A49601	nucleocapsid prote
920	6	1.9	525	2	T48824	nucleocapsid prote
921	6	1.9	525	2	AF3549	hypothetical prote
922	6	1.9	529	2	T35966	multidrug resistan
923	6	1.9	530	1	T44892	probable secreted
924	6	1.9	530	1	T44892	probable secreted
925	6	1.9	530	2	D82412	chaperonin, 60 Kd
926	6	1.9	531	2	T43089	transfer complex p
927	6	1.9	531	2	A98287	glpd gene homolog
928	6	1.9	531	2	A95918	probable oligopept
929	6	1.9	535	2	A46101	protein-tyrosine-p
930	6	1.9	535	2	F90418	ABC transporter, p
931	6	1.9	536	2	B46230	RNA-binding protei
932	6	1.9	538	2	AG3295	phosphoglycerate d
933	6	1.9	540	2	A57624	retinoblastoma pro
934	6	1.9	540	2	T47858	hypothetical prote
935	6	1.9	541	2	AC2353	peptide-chain-rele
936	6	1.9	541	2	T48811	hypothetical prote
937	6	1.9	541	2	A87595	amine oxidase, fla
938	6	1.9	544	2	S41626	spike protein chat
939	6	1.9	546	2	AC2368	hypothetical prote
940	6	1.9	546	2	T19139	hypothetical prote
941	6	1.9	546	2	T02029	DNA-binding protei
942	6	1.9	547	2	B96530	Similar to CCS1 [i
943	6	1.9	548	2	B46101	protein-tyrosine-p
944	6	1.9	548	2	T36881	probable transcrip
945	6	1.9	549	2	AH1019	Sodium,solute symp
946	6	1.9	549	2	S37814	MIF2 protein - yea
947	6	1.9	550	2	E95359	Probable ABC trans
948	6	1.9	550	2	B84484	Mutator-like trans
949	6	1.9	551	2	B82753	dihydrolipoamide S
950	6	1.9	551	2	T22121	hypothetical prote
951	6	1.9	553	1	T02139	calcium-dependent
952	6	1.9	553	2	T45596	telomere repeat-bi
953	6	1.9	554	2	JE0303	propanediol dehydr
954	6	1.9	554	2	A56111	glycerol dehydrata
955	6	1.9	554	2	AC0760	hypothetical prote
956	6	1.9	555	2	T26413	isocitrate lyase (
957	6	1.9	557	1	T28179	hypothetical prote
958	6	1.9	559	2	T22179	5'-nucleotidase (E
959	6	1.9	560	1	HM1V2	hemagglutinin prec
960	6	1.9	562	2	AH2084	hypothetical prote
961	6	1.9	562	2	C71473	hypothetical prote
962	6	1.9	564	2	S36637	signal recognition
963	6	1.9	566	2	T45162	succinate dehydrat
964	6	1.9	568	1	HNN241	hemagglutinin-neur
965	6	1.9	568	2	D90525	hypothetical prote
966	6	1.9	568	2	B86364	hypothetical prote
967	6	1.9	570	2	D145712	Dpp receptor SAX p
968	6	1.9	570	2	D97738	hypothetical prote
969	6	1.9	573	2	AB2401	hypothetical prote
970	6	1.9	574	2	T00245	reverse transcript
971	6	1.9	577	2	T40297	membrane transport
972	6	1.9	578	2	T38775	hypothetical prote
973	6	1.9	579	2	A86851	amino acid permeas
974	6	1.9	579	2	G84583	CDC27/YWC2-like pr
975	6	1.9	580	2	AB0994	gamma-glutamyltran
976	6	1.9	580	2	D86426	hypothetical prote
977	6	1.9	581	1	ERADP5	fiber protein - hu
978	6	1.9	581	2	G83189	probable sodium/hy

```

979 6 1.9 581 2 A4551 insect-stage-speci
980 6 1.9 582 2 F7141 hypothetical prote
981 6 1.9 582 2 A50170 probable transport
982 6 1.9 583 2 F70592 probable lpgB prot
983 6 1.9 587 2 D95247 aspartate-tRNA synt
984 6 1.9 587 2 A98112 aspartate-tRNA lig
985 6 1.9 587 2 A96955 ferrous iron trans
986 6 1.9 589 2 AC1644 ABC transporter (A
987 6 1.9 591 2 S30145 ketol-acid reducto
988 6 1.9 591 2 T45681 ketol-acid reducto
989 6 1.9 592 2 S50972 RIB2 protein - yea
990 6 1.9 592 2 T51712 chreonine ammonia-
991 6 1.9 594 2 S36502 LI protein (altern
992 6 1.9 597 2 C83349 hypothetical prote
993 6 1.9 598 2 D71391 NADH2 dehydrogenas
994 6 1.9 598 2 F83508 flagellar M ring p
995 6 1.9 598 2 D84242 hypothetical prote
996 6 1.9 598 2 T20769 hypothetical prote
997 6 1.9 598 2 AB1236 intercalin protein
998 6 1.9 600 2 AB1331 excludinase ABC c
999 6 1.9 601 2 T11451 NADH2 dehydrogenas
1000 6 1.9 601 2 A55921 serine/threonine k

```

ALIGNMENTS

RESULT 1

conserved hypothetical protein ykud - Bacillus subtilis

A69865
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C/Species: Bacillus subtilis
C/Accession: A69865
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Groullet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
Nature 330, 249-256, 1997
A/Authors: Foulger, D.R.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall
Koeber, P.; Konigstein, G.; Hübner, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Man
Y.M.; Ogawa, K.; Ogiwara, A.; Oudga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scall
A/Authors: Schleich, S.; Schöner, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A/Authors: Yoshikawa, H.F.; Zumbach, E.; Yoshikawa, H.; Zumbach, E.; Yoshikawa, H.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: A69865
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-164 <KUN>
A/Cross-references: GB:29911; GB:AL009126; NID:G2633699; PIDN:CBM13277.1; PTD:6184994;
A/Experimental source: strain 168
C/Genetics:
A/Gene: ykud

Query Match 2.5%; Score 8; DB 2; Length 164;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 YGVKGGT 80
DB 4 YGVKGGT 11

RESULT 2

H84135
hypothetical protein BH388 [Imported] - Bacillus halodurans (strain C-125)
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: H84135

R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and c
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: H84135
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-227 <STO>
A/Cross-references: GB:AP001520; GB:BA000004; NID:G10176401; PIDN:BA07607.1; GSPDB:GN00
A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH388

Query Match 2.5%; Score 8; DB 2; Length 227;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 206 ATNPPVR 213
DB 159 ATNPPVR 166

RESULT 3

S50385
hypothetical protein YLR301W - yeast (Saccharomyces cerevisiae)

N/Alternate names: hypothetical protein L8003.2
C/Species: Saccharomyces cerevisiae
C/Date: 13-Jan-1995 #sequence_revision 20-Feb-1995 #text_change 19-Apr-2002
C/Accession: S50385
R/Pauley, A.
submitted to the EMBL Data Library, November 1994
A/Description: The sequence of S. cerevisiae cosmid 8003.
A/Reference number: S50385
A/Accession: S50385
A/Molecule type: DNA
A/Residues: 1-244 <PAU>
A/Cross-references: EMBL:U17243; NID:G596030; PIDN:AAB67346.1; PID:G596032; GSPDB:GN0001;
C/Genetics:
A/Gene: YLR301W
A/Cross-references: SGD:S0004292
A/Map position: 12R
C/Superfamily: Saccharomyces cerevisiae hypothetical protein YLR301W

Query Match 2.5%; Score 8; DB 2; Length 244;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 221 GSTVTSNG 228
DB 227 GSTVTSNG 234

RESULT 4

T52340
cell wall-plasma membrane linker protein homolog [Imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C/Accession: T52340
R/Guo, J.H.; Park, A.R.; Park, W.J.; Park, O.K.
Plant Mol. Biol. 41, 415-423, 1999
A/Title: Selection of Arabidopsis genes encoding secreted and plasma membrane proteins.
A/Reference number: Z5437; MUID:20064977; PMID:10598107
A/Accession: T52340
A/Status: preliminary; translated from GB/EMBL/DBD
A/Molecule type: mRNA
A/Residues: 1-306 <GOO>
A/Cross-references: EMBL:AF104328; PIDN:AAD11796.1
C/Genetics:
A/Note: CWLP

Query Match 2.5%; Score 8; DB 2; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PRTPPPV 175
 |||||
 Db 140 PRTPPPV 147

RESULT 5

751558
 Probable cell wall-plasma membrane linker protein PRP precursor - rape
 N:Alternate names: hybrid-proline-rich protein
 C:Species: *Brassica napus* (rape)
 C>Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
 C:Accession: S71558
 R:Goodwin, W.; Pallat, J.A.; Jenkins, G.I.
 Plant Mol. Biol. 31, 771-781, 1996
 A>Title: Transcript of a gene encoding a putative cell wall-plasma membrane linker prot
 A:Reference number: S71558; MUID:96400032; PMID:8806408
 A:Accession: S71558
 A:Molecule type: DNA
 A:Residues: 1-376 <COO>
 A:Cross-references: EMBL:X94976; NID:G1155067; PIDN:CAA64425.1; PID:G1155068
 C:Superfamily: hydroxyproline-rich glycoprotein
 P:1-27/Domain: signal sequence #status predicted <SIG>
 P:28-376/Product: probable cell wall-plasma membrane linker protein PRP #status predicted

Query Match 2.5%; Score 8; DB 2; Length 376;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PRTPPPV 175
 |||||
 Db 175 PRTPPPV 182

RESULT 6

S63211
 Cytoskeleton assembly control protein SLA2 - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: MOP2 protein; protein N1102; protein YNL243w; UFG1 protein
 C:Species: *Saccharomyces cerevisiae*
 C>Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 29-Oct-1999
 C:Accession: S63211; S63209; S60673; S61932; S64652; S67355; S72073; S36354; S42278
 R:Saliz, J.E.; Baladron, V.; Del Rey, P.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S63210
 A:Accession: S63211
 A:Molecule type: DNA
 A:Residues: 1-968 <SH1>

A:Cross-references: EMBL:Z71519; NID:G1302279; PIDN:CAA96149.1; PID:e239710; PID:G130228
 A:Experimental source: strain S288C
 R:Pandolfo, D.; De Antoni, A.; Lanfranchi, G.; Valle, G.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S63188
 A:Accession: S63209
 A:Molecule type: DNA
 A:Residues: 820-968 <PAN>

A:Cross-references: EMBL:Z71519
 A:Experimental source: strain S288C
 R:Holtzman, D.A.; Yang, S.; Drudin, D.G.
 J. Cell Biol. 122, 635-644, 1993
 A>Title: Synthetic-lethal interactions identify two novel genes, SLA1 and SLA2, that con

A:Reference number: A40673; MUID:93328765; PMID:8335689
 A:Accession: B40673
 A>Status: nucleic acid sequence not shown
 A:Molecule type: DNA

A:Residues: 1-343, 'A', 345-968 <HO1>
 A:Cross-references: GB:Z22811; NID:G311412; PIDN:CAA80464.1; PID:G311413
 R:Yoon, H.; Donahue, T.F.
 submitted to the EMBL Data Library, March 1994

A:Description: DNA sequence of ufg1 gene in yeast.
 A:Reference number: S61932
 A:Accession: S61932

A:Molecule type: DNA
 A:Residues: 327-343, 'A', 345-559, 'R', 561-886, 'NTVP', 891, 'MLPMHAEVWTTSW', 904 <YOO>

A:Cross-references: EMBL:U07938; NID:G501047; PIDN:AAA19161.1; PID:G501048
 A:Experimental source: strain 117-8A

R:Na, S.; Hincapie, M.; McCusker, J.H.; Haber, J.E.

J. Biol. Chem. 270, 6815-6823, 1995

A>Title: MOP2 (SLA2) affects the abundance of the plasma membrane H(+)-ATPase of *Sacchar*

A:Reference number: S64652; MUID:95204480; PMID:7896828

A:Accession: S64652

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-51, 'A', 53-343, 'A', 345-968 <NAB>

A:Cross-references: EMBL:U12352; NID:G870731; PIDN:AAA74726.1; PID:G870732

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995

R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.

submitted to the EMBL Data Library, February 1996

A:Description: DNA sequence of cosmid 14-5 from chromosome XIV.

A:Reference number: S67355

A:Accession: S67355

A:Molecule type: DNA

A:Residues: 820-968 <PAN>

A:Cross-references: EMBL:Z69381; NID:G1183970; PIDN:CAA93355.1; PID:e221951; PID:G1183971

R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.

Yeast 12, 1071-1076, 1996

A>Title: The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open reading fram

A:Reference number: S72073; MUID:97051596; PMID:8896273

A:Accession: S72073

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 820-968 <PAR>

A:Cross-references: EMBL:Z69381; NID:G1183970; PIDN:CAA93355.1; PID:e221951; PID:G1183971

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996

C:Genetics:

A:Gene: SGD:SLA2; UFG1; END4; MOP2

A:Cross-references: MIPS:YNL243w; SGD:S0005187

A:Map position: 14L

C:Function:

A:Description: essential for the assembly and function of the cortical cytoskeleton

C:Keywords: transmembrane protein

F:771-787/Domain: transmembrane #status predicted <TM>

Query Match 2.5%; Score 8; DB 2; Length 968;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PRTPPPV 175
 |||||
 Db 295 PRTPPPV 302

RESULT 7
 H86885
 50S ribosomal protein L24 [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)

C:Species: *Lactococcus lactis* subsp. *lactis*

C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

C:Accession: H86885

R:Polocin, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich

Gene Res. 11, 731-753, 2001

A>Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ser

A:Reference number: A86625; MUID:2125186; PMID:11337471

A:Accession: H86885

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-101 <STO>

A:Cross-references: GB:AE005176; PID:G12725141; PIDN:AAK06186.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: rplX

C:Superfamily: Escherichia coli ribosomal protein L24

Query Match 2.2%; Score 7; DB 2; Length 101;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 VKTGDIV 285

|||||
Db 3 VKTGDIV 9

RESULT 8

hypothetical protein YMR290w-a - Yeast (Saccharomyces cerevisiae)
C/Species: Saccharomyces cerevisiae
C/Date: 23-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 26-Aug-1999
C/Accession: S69849
R/Barrell, B.G.
submitted to the EMBL Data Library, August 1994
A/Reference number: S47445
A/Accession: S69849
A/Molecule type: DNA
A/Residues: 1-115 <BAR>
A/Cross-references: EMBL:X80836; MIPS:YMR290w-a
C/Genetics:
A/Map position: 13R
C/Superfamily: Saccharomyces hypothetical protein YMR290w-a

Query Match 2.2%; Score 7; DB 2; Length 115;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 PSGALP 304
Db 23 PSCHALP 29

RESULT 9

ribosomal protein L17 - fission yeast (Schizosaccharomyces pombe)
T11639
N/Alternate names: 60S ribosomal protein
C/Species: Schizosaccharomyces pombe
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
R/Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1997
A/Reference number: Z17304
A/Accession: T11639
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-139 <WOO>
A/Cross-references: EMBL:AL021046
R/Lucas, M.; Galliard, C.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A/Reference number: Z21959
A/Accession: T40942
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-139 <LUC>
A/Cross-references: EMBL:AL035259; PIDN:CA22864.1; GSPDB:GN00066; SPDB:SPCC1322.11
A/Experimental source: strain 972h-; cosmid c1322
C/Genetics: <RPL>
A/Map position: 1L
A/Accession: T27153
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-177 <WIL>
A/Cross-references: EMBL:AL032646; PIDN:CA21686.1; GSPDB:GN00019; CESP:Y54E2A
A/Experimental source: clone Y54E2A

Query Match 2.2%; Score 7; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 241 ASNAGTV 247
Db 132 ASNAGTV 138

RESULT 10

T29781
hypothetical protein C50F2.7 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T29781
R/Du, Z.; Le, T.T.
submitted to the EMBL Data Library, November 1996
A/Description: The sequence of C. elegans cosmid C50F2.
A/Reference number: Z20684
A/Accession: T29781
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-141 <DUZ>
A/Cross-references: EMBL:U80445; PIDN:AAB37800.1; GSPDB:GN00019; CESP:C50F2.7
A/Experimental source: strain Bristol N2; clone C50F2
C/Genetics:
A/Gene: CESP:C50F2.7
A/Map position: 1
A/Introns: 19/3; 106/1

Query Match 2.2%; Score 7; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 222 STVTSNG 228
Db 58 STVTSNG 64

RESULT 11

hypothetical protein Y47D3A.20 - Caenorhabditis elegans
T31531
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Aug-2000
C/Accession: T31531
R/McCaw, L.
submitted to the EMBL Data Library, October 1999
A/Reference number: Z21043
A/Accession: T31531
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-161 <WIL>
A/Cross-references: EMBL:AL117202; PIDN:CAB55080.1; CESP:Y47D3A.20
A/Experimental source: clone Y47D3A
C/Genetics:
A/Gene: CESP:Y47D3A.20
A/Introns: 42/1; 55/1; 83/2; 122/1
C/Superfamily: Caenorhabditis elegans hypothetical protein Y47D3A.20

Query Match 2.2%; Score 7; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 GGLAIGS 60
Db 108 GGLAIGS 114

RESULT 12

hypothetical protein Y54E2A.4 - Caenorhabditis elegans
T27153
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C/Accession: T27153
R/Lloyd, C.
submitted to the EMBL Data Library, October 1998
A/Reference number: Z20319
A/Accession: T27153
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-177 <WIL>
A/Cross-references: EMBL:AL032646; PIDN:CA21686.1; GSPDB:GN00019; CESP:Y54E2A
A/Experimental source: clone Y54E2A

C:Genetics:
 A:Gene: CESP:Y54E2A.4
 A:Map position: 1
 A:Insertions: 13/1; 86/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein Y54E2A.4

Query Match 2.2%; Score 7; DB 2; Length 177;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 C:Genetics:
 A:Gene: Xba
 C:Keywords: membrane protein

RESULT 13

AI0213
 Probable membrane protein YP01754 [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AI0213
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tjibbali, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Terraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; Yersinia pestis, PMID:11586360
 A:Accession: AI0213
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-189 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90573.1; PID:g15979781; GSPDB:GN00175
 A:Gene: YP01754
 C:Superfamily: Escherichia coli conserved hypothetical protein b1821

Query Match 2.2%; Score 7; DB 2; Length 189;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 GLRFGVI 25
 |||||
 Db 40 GLRFGVI 46

RESULT 14

JC6024
 kinB sporulation signaling pathway activator kbaa - Bacillus subtilis
 N:Alternate names: integral membrane protein kbaa
 C:Species: Bacillus subtilis
 C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 15-Oct-1999
 C:Accession: JC6024; F69647
 R:Dartois, V.; Djavakhishvili, T.; Hoch, J.A.
 J. Bacteriol. 178, 1178-1186, 1996
 A:Title: Identification of a membrane protein involved in activation of the kinB pathway
 A:Reference number: JC6024; MUID:96165276; PMID:8576055
 A:Accession: JC6024
 A:Molecule type: DNA
 A:Residues: 1-198 <DAR>
 A:Cross-references: GB:U23797; NID:g1143815; PIDN:AA44000.1; PID:g1143817
 R:Kunst, R.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chd A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997
 A:Authors: Poulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Hentut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Konigsstein, G.; Krogh, S.; Kumano, M.; Kurita, A.; Lapidus, A.; Lardinois, A.; Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudaga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelid, R.; Riger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadie, Y.; Sato, T.; Seanton, A.; Authors: Schleif, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, akenchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive Bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: F69647

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-198 <KUN>
 A:Cross-references: GB:299104; GB:AL009126; NID:g2632267; PIDN:CA11932.1; PID:e1182089;
 A:Experimental source: strain 168
 C:Comment: This protein is an intrinsic membrane protein. It is highly hydrophobic and is
 C:Genetics:
 A:Gene: Xba
 C:Keywords: membrane protein

Query Match 2.2%; Score 7; DB 2; Length 198;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LIRFGVI 26
 |||||
 Db 128 LIRFGVI 134

RESULT 15

G86352
 protein T26F17.20 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: G86352
 R:Thoenig, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.; ansen, N.F.; Hughes, B.; Hutzar, L. Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marzilli, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, I. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: G86352
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-208 <STO>
 A:Cross-references: GB:AE005172; NID:g6552738; PIDN:AAFL6537.1; GSPDB:GN00141
 A:Gene: T26F17.20
 A:Map position: 1

Query Match 2.2%; Score 7; DB 2; Length 208;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDEPLTV 320
 |||||
 Db 105 YVDEPLTV 111

RESULT 16

XMECD
 ded protein - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C>Date: 31-Dec-1998 #sequence_revision 17-Oct-1997 #text_change 01-Mar-2002
 C:Accession: H65003; E29803
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: H65003
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-211 <BLAT>

A:Cross-references: GB:AE000320; GB:U00096; NID:G1788647; PIDN:AACT5374.1; PID:G1788653;
A:Experimental source: strain K-12, substrain MG1655
R:Nonet, M.L.; Marvel, C.C.; Tolian, D.R.
J. Biol. Chem. 262, 12209-12217, 1987
A:Title: The hist-purF region of the Escherichia coli K-12 chromosome. Identification of
A:Reference number: A29803; MUID:87308226; PMID:3040734
A:Accession: E29803
A:Molecule type: DNA
A:Residues: 1-221 <NON>
C:Comment: This protein is encoded by a gene located between the hist and purF operons;
C:Genetics:
A:Gene: dedD
A:Map position: 50 min
C:Superfamily: dedD protein

Query Match 2.2%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 178 PAPVAPP 184
|||
Db 92 PAPVAPP 98

RESULT 17
T35071
probable membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
A:Accession: T35071
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Randal, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21567
A:Accession: T35071
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-211 <SE>
A:Cross-references: EMBL:AL096884; PIDN:CA851434.1; GSPDB:GN00070; SCOEDB:SC4G6.11C
C:Genetics:
A:Gene: SCOEDB:SC4G6.11C

Query Match 2.2%; Score 7; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 GTATVAG 221
|||
Db 50 GTATVAG 56

RESULT 18
A69305
conserved hypothetical protein AF0441 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
A:Accession: A69305
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
R.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Godek, A.; Zhou, L.; Oyerbeck, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uteerback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: A69305
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-225 <KLE>
A:Cross-references: GB:AE001074; GB:AE000782; NID:G2689397; PIDN:AAB90793.1; PID:G265018

Query Match 2.2%; Score 7; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 70;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 118 GDLKVR 124
|||
Db 123 GDLKVR 129

RESULT 19
S26440
hypothetical protein 5a - Methanobacterium thermoformicicum plasmid pFV1
C:Species: Methanobacterium thermoformicicum
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 22-Oct-1999
A:Accession: S30305; S26440
R:Neelling, J.; van Keden, F.J.M.; Eggen, R.I.L.; de Vos, W.M.
Nucleic Acids Res. 20, 6501-6507, 1992
A:Title: Modular organization of related Archaeal plasmids encoding different restriction
A:Reference number: S30302; MUID:33126090; PMID:1336177
A:Accession: S30305
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-227 <NOE>
A:Cross-references: EMBL:X68366; NID:G44632; PIDN:CAA48429.1; PID:G44636
C:Genetics:
A:Genome: plasmid

Query Match 2.2%; Score 7; DB 2; Length 227;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 KPTYNST 42
|||
Db 41 KPTYNST 47

RESULT 20
H82435
conserved hypothetical protein VCA0634 [imported] - Vibrio cholerae (strain N16961 serog
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
A:Accession: H82435
R:Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Ginn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: H82435
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <HEI>
A:Cross-references: GB:AB004393; GB:AE003853; NID:G9658042; PIDN:AAF96535.1; GSPDB:GN001;
A:Experimental source: serogroup O1, strain N16961, biotype El Tor
C:Genetics:
A:Gene: VCA0634
A:Map position: 2

Query Match 2.2%; Score 7; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 SKIAPRY 88
|||
Db 192 SKIAPRY 198

RESULT 21
C84340
ABC transport protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
A:Accession: C84340
R:Ng, W.V.; Kennedy, S.P.; Mahlars, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor

Jung, K.H.; Alam, M.; Freiltaa, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of *Haemophilus influenzae* species NRC-1.
A:Reference number: A84160; PMID:20504483; PMID:11016950
A:Accession: C84340
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <STO>
A:Cross-references: GB:AE004437; NID:G10581334; PIDN:AAZ0087.1; GSPDB:GN00138
C:Genetics:
A:Gene: trp2

Query Match
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GTRATVAG 221
|||||
DB 60 GTRATVAG 66

RESULT 22

C70151
hypothetical protein BB0412 - Lyme disease spirochete
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
A:Accession: C70151
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kierlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.
A:Reference number: A70100; PMID:98065943; PMID:9403685
A:Accession: C70151
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-259 <KLE>
A:Cross-references: GB:AE001146; GB:AE000783; NID:G2688312; PIDN:AAC66791.1; PID:G268832
A:Experimental source: strain B31

Query Match
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 NNINSSY 105
|||||
DB 109 NNINSSY 115

RESULT 23

R89455
protein F55A4.4 [imported] - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
A:Accession: R89455
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see website genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: R89455

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <STO>
A:Cross-references: GB:chr_X; PIDN:AA07562.1; PID:G1519659; GSPDB:GN00028; CESP:F55A4.4
C:Genetics:
A:Gene: F55A4.4
A:Map position: X

Query Match
2.2%; Score 7; DB 2; Length 269;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 VDPLTVL 321
|||||
DB 14 VDPLTVL 20

RESULT 24

S57134
hypothetical protein YJR111c - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein J2009
C:Species: *Saccharomyces cerevisiae*
C>Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 19-Apr-2002
A:Accession: S57134
R:Rose, M.; Koetter, P.; Entian, K.D.
Submitted to the Protein Sequence Database, September 1995
A:Reference number: S56848
A:Accession: S57134
A:Molecule type: DNA
A:Residues: 1-283 <ROS>
A:Cross-references: EMBL:Z49611; NID:G1015825; PIDN:CAA9641.1; PID:G1015826; GSPDB:GN00
C:Genetics:
A:Gene: MIPS:YJR111c
A:Cross-references: SGD:S0003872
A:Map position: 10R

Query Match
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 NNINSSY 105
|||||
DB 191 NNINSSY 197

RESULT 25

H98073
hypothetical protein ABC-MSP [imported] - *Streptococcus pneumoniae* (strain R6)
C:Species: *Streptococcus pneumoniae*
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
A:Accession: H98073
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Biasczak, L.; Burgett, S.; Dehoff, B.S.; E
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McHenry, S.; Mc
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
A:Reference number: A97872; PMID:21429245; PMID:11544234
A:Accession: H98073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-284 <RUR>
A:Cross-references: GB:AE007317; PIDN:AAL00421.1; PID:G15459287; GSPDB:GN00174
C:Genetics:
A:Gene: ABC-MSP
C:Superfamily: maltose transport protein malG

Query Match
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 MDGASIV 260
|||||
DB 181 MDGASIV 187

RESULT 26

D69184
UDP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) MTH634 [similarity] - *Methanobac
C:Species: Methanobacterium thermoautotrophicum*
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Jun-2002
A:Accession: D69184

R,Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
 Ki, S.; Church, G.M.; Daniloff, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.;
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: funct
 A:Accession: D69184
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Status: preliminary/ nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-285 <MTM>
 A:Cross-references: GB:AP000844; GB:AE000666; NID:92621707; PIDN:AAB85140.1; PID:9262171
 A:Experimental source: strain Delta H
 A:Genetics: MTH634
 A:Gene: MTH634
 C:Superfamily: Escherichia coli UTP-glucose-1-phosphate uridylyltransferase
 C:Keywords: nucleotidyltransferase

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 285;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 183 PPTVAP 189
 Db 189 PPTVAP 195

RESULT 27
 JH0204
 hypothetical 30.5K protein precursor - *Enterococcus faecalis* plasmid pAM-beta-1
 C:Species: *Enterococcus faecalis*
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 21-Jul-2000
 C:Accession: JH0204
 R:Swinfild, T.J.; Oultam, J.D.; Thompson, D.E.; Brehm, J.K.; Minton, N.P.
 Gene 87, 79-90, 1990
 A:Title: Physical characterisation of the replication region of the *Streptococcus faecalis*
 A:Reference number: F10201; MUID:90236302; PMID:2110101
 A:Accession: JH0204
 A:Molecule type: DNA
 A:Residues: 1-288 <SWT>
 A:Cross-references: EMBL:X17092; NID:93023041; PIDN:ACG8600.1; PID:93023044
 C:Comment: A number of structural features of this protein suggest a localization at the
 C:Genetics:
 A:Genome: plasmid
 C:Superfamily: probable pheromone-responsive protein
 C:Keywords: plasmid replication
 F:1-24/Domain: signal sequence
 F:25-288/Product: hypothetical 30.5K protein #status predicted <MAT>

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 288;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 191 ATGSGV 197
 Db 27 ATGSGV 33

RESULT 28
 S28264
 hydroxyproline-rich glycoprotein - maize
 C:Species: *Zea mays* (maize)
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Sep-1999
 C:Accession: S28264
 R:Raz, R.; Jose, M.; Moysa, A.; Martinez-Izquierdo, J.A.; Pulidomenech, P.
 Mol. Cell. Genet. 233, 252-259, 1992
 A:Title: Different mechanisms generating sequence variability are revealed in distinct
 A:Reference number: S22456; MUID:9223123; PMID:1603067
 A:Accession: S28264
 A:Molecule type: DNA
 A:Residues: 1-303 <RAZ>
 A:Cross-references: EMBL:X63134; NID:922332; PIDN:CAA4844.1; PID:922333
 C:Superfamily: hydroxyproline-rich glycoprotein
 C:Keywords: glycoprotein; hydroxyproline

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 303;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 167 KPTPTP 173
 Db 126 KPTPTP 132

RESULT 29
 AE3325
 cont., competence lipoprotein [imported] - *Brucella melitensis* (strain 16M)
 C:Species: *Brucella melitensis*
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AE3325
 R:Delvecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Ios, T.; Ivanova
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AE3325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-309 <KUR>
 A:Cross-references: GB:AE008917; PIDN:AAL51768.1; PID:917982509; GSPDB:GN00150
 A:Experimental source: strain 16M
 A:Genetics: BME10587
 A:Map position: 1

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 309;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 30 LAGCASK 36
 Db 47 LAGCASK 53

RESULT 30
 G84936
 kinase [imported] - *Buchnera* sp. (strain APS)
 C:Species: *Buchnera* sp.
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C:Accession: G84936
 R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
 Nature 407, 81-86, 2000
 A:Title: Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. A
 A:Reference number: A84930; MUID:20445173; PMID:10993077
 A:Accession: G84936
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-315 <STO>
 A:Cross-references: GB:AP000398; GSPDB:GN00144
 A:Experimental source: strain APS
 A:Genetics: YB3052; BU060

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 315;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 256 GASIVIQ 262
 Db 290 GASIVIQ 296

RESULT 31
 T26886
 hypothetical protein Y44A6.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T26886
 R/Anncough, R.
 Submitted to the EMBL Data Library, January 1998
 A/Reference number: Z20280
 A/Accession: T26886
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-326 <MTL>
 A/Cross-references: EMBL:AL021491, PIDN:CAA16373.1, GSPDB:GN00023, CESP:Y44A6B.1
 A/Experimental source: clone Y44A6B
 C/Genetics:
 A/Gene: CESP:Y44A6B.1
 A/Map position: 5
 A/Intons: 33/1; 128/3; 170/3; 213/1; 265/3

Query Match 2.2%; Score 7; DB 2; Length 326;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 221 GSTVTSN 227
 |||||
 Db 211 GSTVTSN 217

RESULT 32

hydroxyproline-rich glycoprotein precursor - maize

C/Species: Zea mays (maize)
 C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 24-Sep-1999
 C/Accession: JQ0985; PS0215
 R/Stiefel, V.; Ritz-Avila, L.; Raz, R.; Valles, M.F.; Gomez, J.; Pages, M.; Martinez-Izq
 Plant Cell 2, 785-793, 1990
 A/Title: Expression of a maize cell wall hydroxyproline-rich glycoprotein gene in early
 A/Reference number: JQ0985; MUID:93005673; PMID:2152127
 A/Accession: JQ0985
 A/Molecule type: DNA
 A/Residues: 1-328 <STI>
 A/Cross-references: GB:945164; NID:g257040; PIDN:AA23539.1; PID:g257041
 A/Accession: PS0215
 A/Molecule type: protein
 A/Residues: 27-43 <ST2>
 A/Experimental source: leaf
 C/Superfamily: hydroxyproline-rich glycoprotein
 C/Keywords: glycoprotein
 F/1-26/Domain: signal sequence #status predicted <SIG>
 F/27-328/Product: hydroxyproline-rich glycoprotein #status predicted <MAT>

Query Match 2.2%; Score 7; DB 2; Length 328;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 KPTPTTP 173
 |||||
 Db 121 KPTPTTP 127

RESULT 33

probable quinone oxidoreductase - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
 C/Accession: T40981
 R/Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Voicikert, G.
 submitted to the EMBL Data Library, October 1998
 A/Reference number: Z21961
 A/Accession: T40981
 A/Molecule type: DNA
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Residues: 1-329 <LYN>
 A/Cross-references: EMBL:AL031966; PIDN:CAA21450.1; GSPDB:GN00066; SPDB:SPCC1442.16C
 A/Experimental source: strain 972h; cosmId c1442
 C/Genetics:
 A/Gene: SPCC285.01c; SPDB:SPCC1442.16c

A/Map position: 3
 C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 2.2%; Score 7; DB 2; Length 329;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 VKTGDTV 285
 |||||
 Db 138 VKTGDTV 144

RESULT 34

conserved hypothetical protein yyad (replication origin region) - Bacillus subtilis

C/Species: Bacillus subtilis
 C/Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
 C/Accession: I40448; S66018; C70084; S18084
 R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.
 Mol. Microbiol. 6, 629-634, 1992

A/Title: Genes and their organization in the replication origin region of the bacterial
 A/Reference number: I40435; MUID:92204018; PMID:1552862
 A/Accession: I40448
 A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA
 A/Residues: 1-338 <RES>
 A/Cross-references: EMBL:X62539; NID:g40020; PIDN:CAA44412.1; PID:g580907
 R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.
 DNA Res. 1, 1-14, 1994

A/Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom
 A/Reference number: S65967; MUID:96051385; PMID:7584024
 A/Accession: S66018

A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-338 <OGA>

A/Cross-references: EMBL:D26185; NID:g467326; PIDN:BA05224.1; PID:d1005766; PID:g467378
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
 R/Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 C.; Bron, S.; Brouillet, S.; Bruchet, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallert
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
 Koetter, P.; Koningsreйн, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A.; Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maunda, S.; Maneel,
 Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
 A.; Authors: Schleich, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot,
 akenuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A/Authors: Yoshikawa, H.F.; Zamestein, E.; Yoshikawa, H.; Danchin, A.
 A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtlis.
 A/Reference number: A69580; MUID:98044033; PMID:9384337
 A/Accession: C70084

A/Molecule type: DNA
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Accession: C70084
 A/Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CA16131.1; PID:g2636641
 A/Experimental source: strain 168

C/Genetics:
 A/Gene: yyad
 A/Start codon: GNG

Query Match 2.2%; Score 7; DB 2; Length 338;
 Best Local Similarity 100.0%; Pred. No. 16+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 GLIFGYI 25
 |||||
 Db 213 GLIFGYI 219

RESULT 35

T34158
 hypothetical protein C42D4.6 - *Caenorhabditis elegans*
 C/Species: *Caenorhabditis elegans*
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C/Accession: T34158
 R/PU, Z, Le, I.
 submitted to the EMBL Data Library, December 1995
 A/Description: The sequence of C. elegans cosmid C42D4.
 A/Reference number: Z21483
 A/Accession: T34158
 A/Status: preliminary; translated from GB/EMBL/DDBJ
 A/Molecule type: DNA
 A/Residues: 1-339 <DUZ>
 A/Cross-references: EMBL:U41991; PIDN:AAA83342.1; CESP:C42D4.6
 C/Genetics:
 A/Gene: CESP:C42D4.6
 A/Introns: 135/1; 158/3

Query Match 2.2%; Score 7; DB 2; Length 339;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 237 DJLNASN 243
 DB 281 DJLNASN 287

RESULT 36

S19129
 proline-rich protein TRPP-F1 - tomato
 C/Species: *Lycopersicon esculentum* (tomato)
 C/Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 24-Sep-1999
 C/Accession: S19129; S16589; S16751
 R/Salts: Y.; Kenigsbuch, D.; Wachs, R.; Grulsem, W.; Barg, R.
 Plant Mol. Biol. 18, 407-409, 1992
 A/Title: DNA sequence of the tomato fruit expressed proline-rich protein gene TRPP-F1
 A/Reference number: S19129; MUID:92119262; PMID:1731999
 A/Accession: S19129
 A/Molecule type: DNA
 A/Residues: 1-346 <SAL>
 A/Cross-references: EMBL:X61395; NID:G1669697; PIDN:CAA43666.1; PID:G19390
 A/Experimental source: cv. VPNT cherry
 R/Salts: Y.; Wachs, R.; Grulsem, W.; Barg, R.
 Plant Mol. Biol. 17, 149-150, 1991
 A/Title: Sequence coding for a novel proline-rich protein preferentially expressed in y
 A/Reference number: S16589; MUID:91329722; PMID:1868217
 A/Accession: S16589
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 34-346 <SAM>
 A/Cross-references: EMBL:X57076; NID:G19520; PIDN:CAA40361.1; PID:G19521
 A/Experimental source: cv. Arava; haplotype 2n
 C/Genetics:
 A/Gene: TRPP-F1
 A/Map position: 7
 C/Superfamily: hydroxyproline-rich glycoprotein

Query Match 2.2%; Score 7; DB 2; Length 346;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 TPTPPV 175
 DB 190 TPTPPV 196

RESULT 37

S22456
 hydroxyproline-rich glycoprotein - perennial teasinte
 C/Species: *Zea diploperennis* (perennial teasinte)
 C/Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Sep-1999
 C/Accession: S22456; S18964
 R/Raz, R.; Jose, M.; Moya, A.; Martinez-Izquierdo, J.A.; Puigdomenech, F.

Mol. Gen. Genet. 233, 252-259, 1992
 A/Title: Different mechanisms generating sequence variability are revealed in distinct re
 A/Reference number: S22456; MUID:92293123; PMID:1603067
 A/Accession: S22456
 A/Molecule type: DNA
 A/Residues: 1-350 <RAZ>
 A/Cross-references: EMBL:X64173; NID:G22091; PIDN:CAA45514.1; PID:G22092
 C/Superfamily: hydroxyproline-rich glycoprotein
 C/Keywords: glycoprotein; hydroxyproline

Query Match 2.2%; Score 7; DB 2; Length 350;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 KPPTPP 173
 DB 125 KPPTPP 131

RESULT 38

AC0302
 conserved hypothetical protein YPO2474 [imported] - *Yersinia pestis* (strain CO92)
 C/Species: *Yersinia pestis*
 C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C/Accession: AC0302
 R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.;
 deno-Tariga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, E
 Nature 413, 523-527, 2001
 A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A/Reference number: AB0001; MUID:21470413; PMID:11586360
 A/Accession: AC0302
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-351 <YUR>
 A/Cross-references: GB:AL590842; PIDN:CAC91279.1; PID:G15980468; GSPDB:GNO0175
 C/Genetics:
 A/Gene: YPO2474

Query Match 2.2%; Score 7; DB 2; Length 351;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 KRUGLF 22
 DB 176 KRUGLF 182

RESULT 39

G71958
 hypothetical protein jhp0220 - *Helicobacter pylori* (strain J99)
 C/Species: *Helicobacter pylori*
 A/Variety: strain J99
 C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C/Accession: G71958
 R/Alm, R.A.; Ling, L.S.L.; Molt, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Metberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Voyts, G.F.;
 Nature 397, 176-180, 1999
 A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A/Reference number: A71800; MUID:99120557; PMID:9923682
 A/Accession: G71958
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-355 <ARN>
 A/Cross-references: GB:AE001460; GB:AE001439; NID:G4154733; PIDN:AA05807.1; PID:G4154746
 A/Experimental source: strain J99
 C/Genetics:
 A/Gene: jhp0220

Query Match 2.2%; Score 7; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 QGDTVSK 83
 |||||
 Db 287 QGDTVSK 293

RESULT 40

H83202
 poly(beta-d-mannuronate) lyase precursor AlgI PA3547 [imported] - Pseudomonas aeruginosa
 C/Species: Pseudomonas aeruginosa
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: H83202
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A/Reference number: A82950; MUID:2043737; PMID:10984043
 A/Accession: H83202
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-367 <STO>
 A/Cross-references: GB:AE004775; GB:AE004091; NID:9949692; PIDN:AA06935.1; GSPDB:GN001
 A/Experimental source: strain PA01
 C/Genetics:
 A/Gene: algI; PA3547
 C/Superfamily: Pseudomonas aeruginosa poly(beta-D-mannuronate) lyase

Query Match 2.2%; Score 7; DB 2; Length 367;
 Best Local Similarity 100.0%; Freq. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 232 SGRDGL 238
 |||||
 Db 109 SGRDGL 115

Search completed: July 8, 2003, 11:21:37
 Job time : 65 secs

THIS PAGE BLANK (USPTO)


```

983 5 1.6 129 1 KVM HUMAN P18136 homo sapien
984 5 1.6 129 1 KVM MOUSE P01680 mus musculu
985 5 1.6 129 1 KVM ANAPL P00706 anas platyr
986 5 1.6 129 1 KVM CHAM P22310 chrysoloophu
987 5 1.6 129 1 KVM LOPE P22310 chrysoloophu
988 5 1.6 129 1 KVM PAVC P19849 pavo cristat
989 5 1.6 129 1 KVM SYRS P24333 symaticus
990 5 1.6 129 1 KVM THRTN P81711 symaticus
991 5 1.6 129 1 KVM ACAC P08066 acanthamoeb
992 5 1.6 129 1 KVM ACAC P08066 acanthamoeb
993 5 1.6 129 1 KVM ACAC P08066 acanthamoeb
994 5 1.6 129 1 KVM ACAC P08066 acanthamoeb
995 5 1.6 130 1 KVM ACAC P08066 acanthamoeb
996 5 1.6 130 1 KVM ACAC P08066 acanthamoeb
997 5 1.6 131 1 KVM ACAC P08066 acanthamoeb
998 5 1.6 131 1 KVM ACAC P08066 acanthamoeb
999 5 1.6 131 1 KVM ACAC P08066 acanthamoeb
1000 5 1.6 131 1 KVM ACAC P08066 acanthamoeb

```

ALIGNMENTS

RESULT 1
TS1_CNNAL STANDARD; PRT; 196 AA.

```

AC 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thiol-specific antioxidant protein.
GN TSAI.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
[1]
RN SEQUENCE FROM N.A.
RA Marchais V., Cottin J.,
RT "Sequence of TSA-like gene in Candida albicans."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PHYSIOLOGICALLY IMPORTANT ANTIOXIDANT WHICH CONSTITUTES
CC AN ENZYMAIC DEFENSE AGAINST SULFUR-CONTAINING RADICALS. CAN
CC PROVIDE PROTECTION AGAINST A THIOL-CONTAINING OXIDATION SYSTEM BUT
CC NOT AGAINST AN OXIDATION SYSTEM WITHOUT THIOL.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED, UPON OXIDATION.
CC -1- PTM: THE CYS-48-SH GROUP IS THE PRIMARY SITE OF OXIDATION. BY
CC H(2)O(2), AND THE OXIDIZED CYS-48 (PROBABLY CYS-SOH) RAPIDLY
CC REACTS WITH CYS-169-SH OF THE OTHER SUBUNIT TO FORM AN
CC INTERMOLECULAR DISULFIDE. THIS DISULFIDE IS SUBSEQUENTLY REDUCED
CC BY THIOREDOXIN.
CC -1- SIMILARITY: BELONGS TO THE AHP/C-TSA FAMILY.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF149421; AAD34017.1;
CC InterPro: IPR000866; AHP-C-TSA.
CC Pfam: PF00578; AHP-C-TSA; 1.
CC Antioxidant.
CC ACT_SITE 48 48 BY SIMILARITY.
CC ACT_SITE 169 169 BY SIMILARITY.
CC SEQUENCE 196 AA; 21860 MM; 0BD7B6B4715DCDB CRC64;

Query Match 2.5%; Score 8; DB 1; Length 196;
Best Local Similarity 100.0%; Pred.No.1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 PVTGQAP 180
DB 3 PVTGQAP 10

RESULT 2

SLA2_YEAST STANDARD; PRT; 968 AA.

```

AC 01-FEB-1994 (Rel. 28, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SLA2 protein (Transmembrane protein MOP2).
DE SLA2 OR ENDA OR MOP2 OR UFG1 OR YNL243W OR N1102.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RA Holtzman D.A., Yang S., Drubin D.G.,
RT "Synthetic-lethal interactions identify two novel genes, SLA1 and
RT SLA2, that control membrane cytoskeleton assembly in Saccharomyces
RT cerevisiae."
RL J. Cell Biol. 122:635-644 (1993).
[2]
RN SEQUENCE FROM N.A.
RA Holtzman D.A., Yang S., Drubin D.G.,
RT "Synthetic-lethal interactions identify two novel genes, SLA1 and
RT SLA2, that control membrane cytoskeleton assembly in Saccharomyces
RT cerevisiae."
RL J. Cell Biol. 122:635-644 (1993).
[3]
RN SEQUENCE FROM N.A.
RA Holtzman D.A., Yang S., Drubin D.G.,
RT "Synthetic-lethal interactions identify two novel genes, SLA1 and
RT SLA2, that control membrane cytoskeleton assembly in Saccharomyces
RT cerevisiae."
RL J. Cell Biol. 122:635-644 (1993).
[4]
RN SEQUENCE FROM N.A.
RA Holtzman D.A., Yang S., Drubin D.G.,
RT "Synthetic-lethal interactions identify two novel genes, SLA1 and
RT SLA2, that control membrane cytoskeleton assembly in Saccharomyces
RT cerevisiae."
RL J. Cell Biol. 122:635-644 (1993).
[5]
RN SEQUENCE FROM N.A.
RA Holtzman D.A., Yang S., Drubin D.G.,
RT "Synthetic-lethal interactions identify two novel genes, SLA1 and
RT SLA2, that control membrane cytoskeleton assembly in Saccharomyces
RT cerevisiae."
RL J. Cell Biol. 122:635-644 (1993).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Z22811; CAA80464.1;
CC EMBL: L12352; AAA74726.1;
CC EMBL: U07938; AAA19161.1;
CC EMBL: Z71519; CAA96149.1;
CC EMBL: Z71518; CAA96148.1;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z69381; CAA93355.1; -.
 DR PIR; S36354; S36354.
 DR PIR; B40673; B40673.
 DR SGD; S0005187; SLA2.
 DR InterPro; IPR001026; ENTH.
 DR InterPro; IPR002558; ILMEQ.
 DR Pfam; PF01417; ENTH; 1.
 DR Pfam; PF01608; ILMEQ; 1.
 DR ProDom; PD011820; ILMEQ; 1.
 DR SMART; SM00273; ENTH; 1.
 DR SMART; SM00307; ILMEQ; 1.
 DR Structural protein; Cytoskeleton; Actin-binding; Transmembrane.
 KM TRANSMEM 772 791 POTENTIAL.
 FT CONFLICT 52 52 P -> A (IN REF. 2).
 FT CONFLICT 344 344 A -> R (IN REF. 4).
 FT CONFLICT 560 560 A -> R (IN REF. 3).
 FT CONFLICT 887 968 ECKSGYTDKCSRLGNHVKWMIEDHSTSCQOQPLDTSEH
 TLTAEWEEQQVEILKEQSLSNARKLRGELRRHAYNQDDDD
 -> NTVPKMLPMHAEVWVTSV (IN REF. 3).
 SQ SEQUENCE 968 AA; 108910 MW; E592E09D8040CE9 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 968;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PRTPPVV 175
 DB 295 PRTPPVV 302

RESULT 3
 RL23 SCHPO STANDARD; PRT; 139 AA.
 AC 042667;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L23.
 GN (RPL23A OR RPL23 OR SPAC39.03) AND (RPL23B OR SPCC1322.11).
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A. (RPL23A AND RPL23B).
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris P., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
 RA Welfens I., Vansireels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moeserl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Motiller S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucan M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong U., Forzburg S.L.,
 RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RA "The genome sequence of Schizosaccharomyces pombe";

RL Nature 415:871-880(2002).
 CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L23 IN S. POMBE.
 CC -1- SIMILARITY: BELONGS TO THE L1AP FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL021046; CAA15912.1; -.
 DR EMBL; AL035259; CAA22864.1; -.
 DR HSP; P04450; 1WH1.
 DR InterPro; IPR000218; Ribosomal_L14.
 DR Pfam; PF00238; Ribosomal_L14; 1.
 DR ProDom; PD001093; Ribosomal_L14; 1.
 DR PROSITE; PS00049; RIBOSOMAL_L14; 1.
 KM Ribosomal protein; Multigene family.
 SQ SEQUENCE 139 AA; 14882 MW; D6C9A8903225FAD CRC64;

Query Match 2.2%; Score 7; DB 1; Length 139;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 ASNAGTV 247
 DB 132 ASNAGTV 138

RESULT 4
 YCF7_EUGER STANDARD; PRT; 161 AA.
 AC P48337;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Hypothetical 18.6 kDa protein ycf7 (ORF161).
 GN YCF7.
 OS Euglena gracilis.
 OG Chloroplast.
 OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
 OX NCBI_TaxID=3039;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2;
 RX MEDLINE=93347989; PubMed=8346031;
 RA Hallick R.B., Hong L., Drager R.G., Favreau M.R., Monfort A.,
 RA Great B., Spielmann A., Stutz E.,
 RT "Complete sequence of Euglena gracilis chloroplast DNA."
 RL Nucleic Acids Res. 21:3537-3544(1993).
 CC -1- SIMILARITY: BELONGS TO THE YCF7 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z11874; -; NOT ANNOTATED_CDS.
 DR EMBL; X70810; CAA50117.1; -.
 DR InterPro; IPR001700; RNA_pol_A_bac.
 DR Pfam; PF01000; RNA_pol_A_bac; 1.
 KM Chloroplast; Hypothetical protein.
 SQ SEQUENCE 161 AA; 18606 MW; 377944A5C814949C CRC64;

Query Match 2.2%; Score 7; DB 1; Length 161;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 OKPIKRL 18
Db 110 OKPIKRL 116

RESULT 5

KBAA_BACSU STANDARD, PRT, 198 AA.

AC Pl643; 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE KINB signaling pathway activation protein.

OS KBAA.

OC Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

OC NCBI_TaxID=1423;

RA Yon J.R., Sammons R.L., Smith D.A.;
RT "Cloning and sequencing of the gerD gene of Bacillus subtilis.";
RL J. Gen. Microbiol. 135:3431-3445(1989).
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION OF THE KINB SIGNALING PATHWAY
CC OF SPOULATION
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; U23797; AAC44000.1; -;
CC EMBL; D64126; BAA10996.1; -;
CC EMBL; Z99104; CAB11932.1; -;
CC EMBL; M27259; AAA22469.2; ALT INIT.
CC Subtilisin; Bg10643; KbaA.
CC Sporulation; Transmembrane; Complete proteome.
CC TRANSMEM 9 29 POTENTIAL.
CC TRANSMEM 42 62 POTENTIAL.
CC TRANSMEM 90 110 POTENTIAL.
CC TRANSMEM 117 137 POTENTIAL.
CC TRANSMEM 146 166 POTENTIAL.
CC TRANSMEM 173 193 POTENTIAL.
CC SEQUENCE 198 AA; 22775 MW; 24917B51F38E0ACD CRC64;

Query Match 2.2%; Score 7; DB 1; Length 198;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 20 LIRGVIT 26
Db 128 LIRGVIT 134

RESULT 6
DEDD_ECOLI STANDARD, PRT, 220 AA.

AC P09549; 01-MAR-1989 (Rel. 10, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DEDD protein.
GN DEDD OR B2314.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;

RA NINET M.L., Mareel C.C., Tolan D.R.;
RT "The hnf-purF region of the Escherichia coli K-12 chromosome."
RT Identification of additional genes of the hnf and purF operons."
RL J. Biol. Chem. 262:12209-12217(1987).

RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RL [3]
RA SEQUENCE FROM N.A.
RA STRAIN=K12;
RA MEDLINE=97349980; PubMed=9205837;

RA SEQUENCE FROM N.A.
RA STRAIN=K12;
RA MEDLINE=97349980; PubMed=9205837;

Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakada S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Satoh N., Sampei G., Sato Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horinouchi T.,
 RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
 RT K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.",
 RL DNA Res. 4:91-113(1997).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; M68934; AAA23967.1; ALT INIT.
 DR EMBL; AE000320; AAC75374.1; ALT INIT.
 DR EMBL; D90862; BAA16162.1; ALT INIT.
 DR EMBL; D90863; BAA16170.1; ALT INIT.
 DR PIR; E29803; XMECD.
 DR Ecogene; EG10218; dedb.
 KW Complete proteome.
 SQ SEQUENCE 220 AA; 22938 MW; 454529DCB8AC1644 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 220;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 PAPVAP 184
 DB 101 PAPVAP 107

RESULT 7

YPVA_METTF STANDARD; PRT; 227 AA.

AC P29577;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Hypothetical 24.7 kDa protein (ORF5A).
 OS Methanobacterium thermoformicicum.
 OG Plasmid pPV1.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 NCBI_TaxID=145262;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 3848 / THF;
 RX MEDLINE=93126090; PubMed=1336177;
 RA Noelling J., Van Breden F.J.M., Eggen R.I.L., de Vos W.M.;
 RT "Molecular organization of related Archaeal plasmids encoding different
 RT restriction-modification systems in Methanobacterium
 RT thermoformicicum.",
 RL Nucleic Acids Res. 20:6501-6507(1992).
 CC -1- SIMILARITY: TO ORFS IN PPZ1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X68366; CAA8429.1; --
 DR PIR; S26440; S26440.
 DR PIR; S30305; S30305.
 DR InterPro; IPR001434; DUF11.

DR Pfam; PF01345; DUF11.1.
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 227 AA; 24681 MW; 2F3004C2519E849E CRC64;

Query Match 2.2%; Score 7; DB 1; Length 227;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 KPTYNST 42
 DB 41 KPTYNST 47

RESULT 8

YJ81_YEAST STANDARD; PRT; 283 AA.

AC P47148;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 32.2 kDa protein in CPA2-NMF1 intergenic region.
 OS YJR111C OR J2009.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Rose M., Koester P., Ertan K.D.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z49611; CAA89641.1; --
 DR SCD; S0003872; YJR111C.
 KW Hypothetical protein.
 SQ SEQUENCE 283 AA; 32208 MW; CD49258DE05D384D CRC64;

Query Match 2.2%; Score 7; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 NNINSSY 105
 DB 191 NNINSSY 197

RESULT 9

T212_LACLC STANDARD; PRT; 304 AA.

AC P50177;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Type II restriction enzyme LladCHI (EC 3.1.21.4) (Endonuclease
 DE LladCHI) (R.LladCHI) (LlaII).
 GN LladCHI OR LlaII.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 NCBI_TaxID=1359;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=DCH-4;
 RX MEDLINE=95314272; PubMed=7793939;
 RA Moineau S., Walker S.A., Vedamuthu E.R., Vandenberg P.A.;
 RT "Cloning and sequencing of LladCHI restriction/modification genes

RT from *Lactococcus lactis* and relatedness of this system to the
RT *Streptococcus pneumoniae* DpnII system.";
RL Appl. Environ. Microbiol. 61:2193-2202(1995).
CC -1- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED UNMETHYLATED SEQUENCE
CC GATC AND CLEAVES BEFORE G-1.
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
CC specific double-stranded fragments with terminal 5'-phosphates.
CC -1- SIMILARITY: TO R.DNII AND R.MBOI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U16027; AAB06313.1; -.
CC DR REBASE; 2772; LIADCHI.
CC KM Hydrolyase; Endonuclease; Nuclease; Restriction system; Plasmid.
SQ SEQUENCE 304 AA; 34670 MW; C0C84547A28A20E4 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 TVSKIAQ 86
Db 171 TVSKIAQ 177

RESULT 10
ID Y060_BUCAI STANDARD; PRT; 315 AA.
AC P57168;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical sugar kinase BU060.
GN BU060.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- SIMILARITY: BELONGS TO THE PKB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001118; BAB12783.1; -.
CC DR InterPro; IPR002173; PfkB.
DR Pfam; PF00294; PfkB; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; FALSE NEG.
KW Hypothetical protein; Transferase; Kinase; Complete proteome.
SQ SEQUENCE 315 AA; 35023 MW; 69B3707C601EFD25 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 256 GASTVIO 262
Db 290 GASTVIO 296

RESULT 11
ID YVAD_BACSU STANDARD; PRT; 338 AA.
AC P37520;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein YVAD.
GN YVAD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / CRK2000;
RX MEDLINE=92204018; PubMed=1552862;
RA Ogasawara N., Yoshikawa H.;
RT "Genes and their organization in the replication origin region of the
RT bacterial chromosome.";
RL Mol. Microbiol. 6:629-634(1992).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=168;
CC MEDLINE=98044033; PubMed=9384377;
CC Kunz F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
CC Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borchert S.,
CC Boursier L., Boursier L., Brans A., Braun M., Brisset S.C., Bron S.,
CC Brouillet S., Bruschi C.V., Caldwell I.F., Cummings N.M., Daniel R.A.,
CC Choi S.K., Codani J.J., Conerton I.F., Emmling S.D., Emerson P.T.,
CC Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Foulger D.,
CC Entian K.D., Errington J., Fabret C., Ferrari E., Galleron N.,
CC Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Gallon N.,
CC Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
CC Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
CC Hilbert H., Holappell S., Hosono S., Hullo M.F., Itaya M., Jones L.,
CC Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
CC Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
CC Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
CC Lee S.W., Levine A., Liu H., Maeda S., Manuel C., Medigue M.,
CC Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
CC Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudaga B., Park S.H.,
CC Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
CC Pressac E., Pujic P., Puttelne B., Rapoport G., Rey M., Reynolds S.,
CC Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadis Y.,
CC Sato I., Scallan E., Schleich S., Schroeter R., Scottone G.,
CC Setlow R., Taconi E., Takai T., Takahashi H., Takemaru K.,
CC Setlow R., Taconi E., Takai T., Takahashi H., Takemaru K.,
CC Tokochi A., Tsuchioka A., Tanaka T., Teruya F., Tognoni A.,
CC Toseo V., Uchiyama S., Vandenbol M., Vannier P., Vassacotti A.,
CC Varti A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
CC Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
CC Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT subtilis.";
RL Nature 390:249-256(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62539; CAA44412.1; -
DR EMBL; D26185; BAA05224.1; -
DR EMBL; Z99124; CAB16131.1; -
DR PIR; S18084; S18084.
DR Subtilist; BG10052; VYAD.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 338 AA; 37721 MW; 8E5DE922917BBE84 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 338;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 GLIFGVY 25
DB 213 GLIFGVY 219

RESULT 12
PRF1_LYCES STANDARD; PRT; 346 AA.
AC 000451;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 36.4 kDa proline-rich protein.
GN PRP-F1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. VENT Cherry; TISSUE=Fruit;
RX MEDLINE=92119262; PubMed=1731999;
RA Salts Y., Wachs R., Kenigsbuch D., Gruissem W., Barg R.;
RT DNA sequence of the tomato fruit expressed proline-rich protein gene
RT PRP-F1 reveals an intron within the 3 untranslated transcript.";
RL Plant Mol. Biol. 18:407-409(1992).
RN [2]
RP SEQUENCE OF 34-346 FROM N.A.
RC STRAIN=cv. Arava;
RX MEDLINE=91329722; PubMed=1868217;
RA Salts Y., Wachs R., Gruissem W., Barg R.;
RT "Sequence coding for a novel proline-rich protein preferentially
RT expressed in young tomato fruit.";
RL Plant Mol. Biol. 17:149-150(1991).
RN [3]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61395; CAA43666.1; -
DR EMBL; X57076; CAA40361.1; -
DR HSSP; P24337; 1HYP.
DR InterPro; IPR001768; Try/amy1_inhbt.
DR Pfam; PF00234; Tryp_alpha_amy1.1.
SQ SEQUENCE 346 AA; 36375 MW; 604E58452F8F16E0 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 346;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 TPTPPVV 175

```

```

DB 190 TPTPPVV 196

RESULT 13
ALGL_PSEAE STANDARD; PRT; 367 AA.
AC 006749; Q57292;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alginatase lyase precursor (EC 4.2.2.3) (Poly(beta-D-mannuronate) lyase)
DE (Poly(mana) alginatase lyase).
GN ALGL OR PA3547.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93380656; PubMed=8370530;
RA Boyd A., Ghosh M., May T.B., Shinabarger D., Keogh R.,
RA Chakrabarty A.M.;
RT "Sequence of the algl gene of Pseudomonas aeruginosa and purification
RT of its alginatase lyase product.";
RL Gene 131:1-8(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRD1;
RX MEDLINE=93328683; PubMed=8335634;
RA Schiller N.L., Monday S.R., Boyd C.M., Keen N.T., Ohman D.E.;
RT "Characterization of the Pseudomonas aeruginosa alginatase lyase gene
RT (algl): cloning, sequencing, and expression in Escherichia coli.";
RL J. Bacteriol. 175:4780-4783(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.R.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [4]
CC -1- FUNCTION: Depolymerizes alginate by cleaving the beta-1,4
CC glycosidic bond. May enhance the production of alginate by
CC controlling the length of the polymer chain during export.
CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of polysaccharides
CC containing beta-D-mannuronate residues to give oligosaccharides
CC with 4-deoxy-alpha-L-erythro-hex-4-enopyranuronosyl groups at
CC their ends.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L14597; AAA71990.1; -
DR EMBL; U27829; AAA91127.1; -
DR EMBL; AE004775; AAG06935.1; -
DR PIR; JN0777; JN0777.
DR Lyase; Signal; Periplasmic; Complete proteome.
DR SIGNAL 1 27
FT CHAIN 28 367
FT CONFLICT 269 269 A -> P (IN REF. 1).

```


FT CONFLICT 337 341 KMLEA -> NACSNRP (IN REF. 1).
 SO SEQUENCE 367 AA; 40829 MW; EAA3FE30032AB3BA CRC64;
 Query Match 2.2%; Score 7; DB 1; Length 367;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 232 SGRDGL 238
 Db 109 SGRDGL 115

RESULT 14
 ID -RL3_YEAST STANDARD; PRT; 386 AA.
 AC P14126; Q08459; Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 37, Last annotation update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 60S ribosomal protein L3 (YL1) (RPL1) (Trichodermin resistance protein).
 OS RPL3 OR TCM1 OR MAK8 OR YOR063W OR YOR29-14.
 OC Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLP1;
 RX MEDLINE=83238226; PubMed=6305925;
 RA Schulerz L.D., Friesen J.D.;
 RT "Nucleotide sequence of the tcm1 gene (ribosomal protein L3) of
 RL Saccharomyces cerevisiae."
 RI J. Bacteriol. 155:8-14(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97279235; PubMed=9133743;
 RA Valens M., Bohn C., Dagnan-Fornier B., Dang V., Bolotin-Fukuhara M.;
 RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals
 RL the presence of two tRNAs and 24 new open reading frames."
 RN Yeast 13:379-390(1997).
 RN [3]
 RP SEQUENCE OF 1-19.
 RX MEDLINE=84038947; PubMed=6355773;
 RA Oetka E., Higo K.I., Itoh T.;
 RT "Yeast ribosomal proteins: VII. Cytoplasmic ribosomal proteins from
 RL Schizosaccharomyces pombe."
 RN Mol. Genet. 191:519-524(1983).
 RN [4]
 RP SEQUENCE OF 1-9.
 RX MEDLINE=92184799; PubMed=1544921;
 RA Takakura H., Tsunashima S., Miyagi M., Warner J.R.;
 RT "NH2-terminal acetylation of ribosomal proteins of Saccharomyces
 RL cerevisiae."
 RN J. Biol. Chem. 267:5442-5445(1992).
 CC -1- FUNCTION: THE L3 PROTEIN IS A COMPONENT OF THE LARGE SUBUNIT OF
 CC CYTOPLASMIC RIBOSOMES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- MISCELLANEOUS: A MUTANT ALLELE OF TCM1 CONFERS RESISTANCE TO
 CC TRICHODERMIN, A TRICHOCECENE TOXIN PRODUCED BY PLANT-PATHOGENIC
 CC FUNGI.
 CC -1- SIMILARITY: BELONGS TO THE LBP FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J01351; AAA88732.1; -
 CC DR EMBL; Z74971; CAA9256.1; -

DR EMBL; Z70678; CAA94548.1; -
 DR PIR; S05843; RSBYAE.
 DR SGD; S0005589; RPL3.
 DR InterPro; IPR000597; Ribosomal_L3.
 DR Pfam; PF00297; Ribosomal_L3; 1.
 DR PROSITE; PS00474; RIBOSOMAL_L3; 1.
 KW Ribosomal protein.
 FT INIT MET 0
 FT CONFLICT 254 254
 FT SEQUENCE 366 AA; 43626 MW; 8202CA08C071CB CRC64;
 SO SEQUENCE

Query Match 2.2%; Score 7; DB 1; Length 386;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 159 TPVVVV 165
 Db 80 TPVVVV 86

RESULT 15
 ID -ALP_TRIHA STANDARD; PRT; 409 AA.
 AC O03420;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Alkaline proteinase precursor (EC 3.4.21.-) (ALP).
 GN PRP1.
 OS Trichoderma harzianum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Microsporici Hypocreales; Trichoderma.
 NX NCBI_TaxId=5544;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=IMI 206040;
 RX MEDLINE=93316857; PubMed=8326868;
 RA Genenta R.A., Goldman G.H., Jacobs D., Ardiles W., Vila S.B.,
 RA Montagu M., Herrera-Betrelia A.;
 RT "Molecular characterization of the proteinase-encoding gene, prp1,
 RL related to mycoparasitism by Trichoderma harzianum."
 RN Mol. Microbiol. 8:603-613(1993).
 CC -1- FUNCTION: SERINE PROTEASE. SECRETED SPECIFICALLY DURING THE
 CC MYCOPARASITIC PROCESS, WHICH IS INVOLVED IN THE DEGRADATION OF
 CC PHYTOPATHOGEN CELL WALLS, MEMBRANES AND OF THE PROTEINS RELEASED
 CC AFTER LYSIS OF THE HOST.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- INDUCTION: BY MYCECIN, FUNGAL CELL WALLS, AND CHITIN, BUT ONLY
 CC IN THE ABSENCE OF GLUCOSE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M87518; AAA34211.1; -
 CC DR EMBL; M87516; AAA34209.1; -
 CC PIR; S32905; S32905.
 CC HSSP; O99405; IMPT.
 CC MEROPS; S08.00A;
 CC InterPro; IPR000209; Peptidase_S8.
 CC Pfam; PF00082; Peptidase_S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC PROSITE; PS00136; SUBTILASE_ASP; 1.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SBR; 1.
 KW Hydrolyase; Serine protease; Zymogen; Signal.
 FT SIGNAL 1
 FT PROPEP 21 120
 FT POTENTIAL.

FT CHAIN 121 409 ALKALINE PHOSPHATASE.
 FT ACT SITE 161 161 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 192 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 409 AA; 42271 MW; D0F548EA236C5B63 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 409;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 SHRTSGS 53
 Db 136 SHRTSGS 142

RESULT 16
 GAT3_BRARE STANDARD; PRT; 438 AA.
 ID GAT3_BRARE
 AC 091428;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor GATA-3 (GATA Binding Factor-3).
 GN GATA3 OR GATA3.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96028520; Pubmed=7547465;
 RA Neave B., Rodaway A., Wilson S.W., Patient R., Holder N.;
 RT "Expression of zebrafish GATA 3 (gata3) during gastrulation and
 Mech. Dev. 51:169-182(1995)."
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING GASTRULATION IN THE VENTRAL
 CC REGION OF THE EMBRYO WHICH INCLUDES TISSUE PATED TO FORM THE NON-
 CC NEURAL ECTODERM.
 CC -1- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, S80425; AAA93491.1; -.
 CC HSSP; P17679; 1GNF.
 CC DR ZFIN; ZDB-GENE-990415-82; gata3.
 CC DR InterPro; IPR000679; Znf.GATA.
 CC DR InterPro; IPR001164; hRIF_1like.
 CC DR Pfam; PF00320; GATA; 2.
 CC DR PRINTS; PRO0619; GATAZNFINGER.
 CC DR SMART; SM00401; Znf.GATA; 2.
 CC DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
 CC DR PROSITE; PS00114; GATA_ZN_FINGER_2; 2.
 CC KM Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein.
 FT ZN_FING 256 280 GATA-TYPE 1.
 FT ZN_FING 310 334 GATA-TYPE 2.
 FT DOMAIN 69 72 POLY-PRO.
 FT DOMAIN 130 135 POLY-SER.
 FT DOMAIN 235 238 POLY-SER.
 FT DOMAIN 315 319 POLY-THR.
 SQ SEQUENCE 438 AA; 47590 MW; 528C929C9775653C CRC64;

Query Match 2.2%; Score 7; DB 1; Length 438;
 Best Local Similarity 100.0%; Pred. No. 38;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 41 STSGSGS 47
 Db 161 STSGSGS 167

RESULT 17
 GATA_MYCPN STANDARD; PRT; 478 AA.
 ID GATA_MYCPN
 AC P75534;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT
 DE subunit A).
 GN GATA OR MEN237 OR MP594.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_Taxid=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; Pubmed=8948633;
 RA Hiemmelreich R., Hilbert H., Plagens H., Pirk1 E., Li B.-C.,
 RA Hermann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae." Res. 24:4420-4449(1996).
 RL Nucleic Acids.
 CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
 CC GIN-tRNA (Gln) THROUGH THE TRANSMUTATION OF MISACTYLATED GLU-
 CC tRNA (Gln) IN ORGANISMS WHICH LACK GLUTAMINYL-tRNA SYNTHETASE. THE
 CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
 CC AN ACTIVATED GAMMA-PHOSPHO-GLU-tRNA (Gln) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA (Gln) + L-glutamine = ADP
 CC + phosphate + L-glutamyl-tRNA (Gln) + L-glutamate.
 CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000058; AAB96242.1; -.
 CC DR InterPro; IPR000120; Amidase.
 CC DR InterPro; IPR004412; Gata.
 CC DR Pfam; PF01425; Amidase; 1.
 CC DR TIGRFAMs; TIGR00132; gata; 1.
 CC DR PROSITE; PS00571; AMIDASSS; 1.
 CC KM Protein biosynthesis; Ligase; Complete proteome.
 SQ SEQUENCE 478 AA; 53229 MW; F228E9A84F641071 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 ISRGVY 314
 Db 187 ISRGVY 193

RESULT 18
 KIR2_HUMAN STANDARD; PRT; 505 AA.
 ID KIR2_HUMAN
 AC P36856; Q15479; Q15480; Q15481; Q15482;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine/threonine-protein kinase receptor R2 precursor (EC 2.7.1.37)

DE (SKR2) (Activin receptor-like kinase 4) (ALK-4) (ACTR-1B).
 DR ACVR1B OR ACVR1K4.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=9330967; PubMed=8397373;
 RA Ten Dijke P., Ichijo H., Franzen P., Schulz P., Sarra J.,
 RT Toyoshima H., Heldin C.-H., Miyazono K.;
 RT "Activin receptor-like kinases: a novel subclass of cell-surface
 RT receptors with predicted serine/threonine kinase activity.";
 RL Oncogene 8:2879-2887(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=94254839; PubMed=8196624;
 RA Carcamo J., Weiss F.M., Ventura F., Wieser R., Wrana J.L.,
 RT Attisano L., Massague J.;
 RT "Type I receptors specify growth-inhibitory and transcriptional
 RT responses to transforming growth factor beta and activin.";
 RL Mol. Cell. Biol. 14:3810-3821(1994).
 RN [3]
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RC TISSUE=Liver;
 RX MEDLINE=94336666; PubMed=8058741;
 RA Xu J., Matsuzaki K., McKeehan K., Wang F., Kan M., McKeehan W.L.;
 RT "Genomic structure and cloned cDNAs predict that four variants in the
 RT kinase domain of serine/threonine kinase receptors arise by
 RT alternative splicing and poly(A) addition.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7957-7961(1994).
 RP [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1/SKR2-1).
 RC TISSUE=Eye;
 RX Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RA Straubeberg R.;
 RT "-1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
 CC "-1- SUBCELLULAR LOCATION: Type 1 membrane protein.
 CC "-1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1/SKR2-1 (SHOWN HERE), 2/SKR2-2
 CC AND 3/SKR2-3; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC "-1- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. MOST STRONGLY IN
 CC HUMAN KIDNEY, PANCREAS, BRAIN, LUNG, AND LIVER.
 CC "-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC TGF-RECEPTOR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL: Z22536; CAAB0258.1; -
 CC EMBL: U14722; AA50246.1; -
 CC EMBL: L10125; AA50555.1; -
 CC EMBL: L10126; AA50556.1; -
 CC EMBL: L31846; AA53349.1; -
 CC EMBL: L31848; AA53350.1; -
 CC EMBL: L31848; AA53351.1; -
 CC EMBL: BC000254; AA00254.1; -
 CC EMBL: L31848; S37184.
 CC EMBL: L31848; S37184.
 CC HSP: P36897; ITBI.
 CC Genew: HSNCL172; ACVR1B.
 CC MIM: 601300; -
 CC InterPro: IPR000472; Activin_rec.
 CC InterPro: IPR000719; Euk_pkinase.
 CC InterPro: IPR002290; Ser_thr_kinase.
 CC InterPro: IPR003605; TGFbeta_GS.
 CC Pfam: PF00069; Pkinase; 1.
 CC Pfam: PF01064; Activin_rec; 1.
 DR

DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00467; GS; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR Receptor; Transferrin; Serine/threonine-protein kinase; ATP-binding;
 KW Transmembrane; Glycoprotein; Signal; Alternative splicing;
 KW Polymorphism.
 FT CHAIN 1 23
 FT STGM 24 505
 FT DOMAIN 24 126
 FT TRANSMEM 127 146
 FT DOMAIN 150 505
 FT DOMAIN 207 497
 FT BINDING 213 221
 FT ACT_SITE 234 335
 FT CARBOHYD 43 43
 FT VARSPPLIC 422 505
 FT
 FT VARSPPLIC 465 505
 FT
 FT VARIANT 408 408
 FT
 FT CONFLICT 56 56
 FT CONFLICT 222 223
 FT SEQUENCE 505 AA; 56806 MW; 40AC65CAAC7573 CRC64;
 SQ
 Query Match 2.2%; Score 7; DB 1; Length 505;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 187 STSGSGS 193
 QY 41 STSGSGS 47
 DB 187 STSGSGS 193
 RESULT 19
 KIR2_RAT STANDARD; PRT; 505 AA.
 AC P80202;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine/threonine-protein kinase receptor R2 precursor (EC 2.7.1.37)
 DE (SKR2).
 GN ACVR1B OR ACVR1K4.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=urogenital ridge;
 RX MEDLINE=9337378; PubMed=8395914;
 RA He W.-W., Gustafson M., Hirobe S., Donahoe P.;
 RT "Developmental expression of four novel serine/threonine kinase
 RT receptors homologous to the activin/transforming growth factor-beta
 RT type II receptor family.";
 RL Dev. Dyn. 196:133-142(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95113007; PubMed=7813622;
 RA Takami T., Moustakas A., Lin H.Y., Lodish H.F.;
 RT "Molecular characterization of a type I serine-threonine kinase
 RT receptor for TGF-beta and activin in the rat pituitary tumor cell
 RT line GH3.";
 RL Exp. Cell Res. 216:208-214(1995).
 RT

CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: UROGENITAL RIDGE, TESTIS, OVARY, BRAIN AND LUNGS.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. TGRB RECEPTOR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S76466; AAB3045.1; -.
 DR HSSP: P36997; 1TBT.
 DR InterPro: IPR000472; Activin_rec.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR003605; TGFbeta_GS.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01064; Activin_rec; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00467; GS; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00108; PROTEIN KINASE SER; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Receptor; Transferrase; Serine/threonine-protein kinase; ATP-binding; Transmembrane; Glycoprotein; Signal.
 FT CHAIN 1 23
 FT SIGNAL 1 23
 FT SERINE/THREONINE-PROTEIN KINASE RECEPTOR
 FT R2.
 FT EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 127 126
 FT TRANSMEM 127 149
 FT DOMAIN 127 149
 FT NP BIND 207 497
 FT BINDING 213 221
 FT ACT SITE 234 234
 FT CARBOHYD 335 335
 FT SEQUENCE 43 43
 SQ SEQUENCE 505 AA; 56804 MW; 377FPA5C867B3860 CRC64;
 Query Match 2.2%; Score 7; DB 1; Length 505;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 STSGSGS 47
 DB 187 STSGSGS 193
 RESULT 20
 EXU_DROME STANDARD; PRT; 532 AA.
 AC P28750; Q9V967;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 15-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-UTN-2002 (Rel. 41, Last annotation update)
 DE Maternal exuperantia protein.
 GN EXU OR CG8994.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_TaxID=7227;
 RN NCB1
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92097546; PubMed=1756733;
 RA Marcey D., Watkins W.S., Hazelrigg T.;
 RT "The temporal and spatial distribution pattern of maternal exuperantia protein: evidence for a role in establishment but not maintenance of bicoid mRNA localization.";
 RT EMBO J. 10:4259-4266(1991).

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9209725; PubMed=1752438;
 RA MacDonald P.M., Luk S.K.-S., Kilpatrick M.;
 RT "Protein encoded by the exuperantia gene is concentrated at sites of bicoid mRNA accumulation in Drosophila nurse cells but not in oocytes or embryos.";
 RT Genes Dev. 5:2455-2466(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gliokh A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostali M., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush F., Kaplen G.H., Ke Z., Kienison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrel A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Snie B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: ENSURES THE PROPER LOCALIZATION OF THE MRNA OF THE BICOID GENE TO THE ANTERIOR REGIONS OF THE OOCYTE THUS PLAYING A FUNDAMENTAL ROLE IN THE ESTABLISHMENT OF THE POLARITY OF THE OOCYTE. MAY BIND THE BCD MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S72757; AAB20673.1; -.
 DR EMBL: S72363; AAB20670.1; -.
 DR EMBL: AB003791; AAB68399.1; -.
 DR PIR: S18643; S18643.
 DR PIR: A11639; A11639.
 DR FlyBase: FBgn000615; exu.
 KW Developmental protein; RNA-binding.
 FT DOMAIN 207 246
 FT VARIANT 223 223
 FT VARIANT 339 339
 R -> S (IN P342 MUTANT, LOSS OF EXU)

FT FUNCTION IN FEMALE.
K -> Q (IN REF. 1).
SQ SEQUENCE 176 176 532 AA; 57974 MW; 50BD15B712A62C4E CRC64;
Query Match 2.2%; Score 7; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
Db 209 STSGSGS 215

RESULT 21
CUEO_YERPE STANDARD; PRT; 533 AA.
AC Q2ERKO; 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Blue copper oxidase cueo precursor (Copper efflux oxidase).
GN CUEO OR YPO3409.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
ON NCBI_TaxID=632;
RX STRAIN=CO-92 / Bivovar Orientalis;
RA MEDLINE=21470413; PubMed=11586360;
RA Parikhil J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellingwell T., Hamlin N., Holroyd S., Jagels K., Kariyehov A.V.,
RA Leather S., Mouton S., Oyston P.C.F., Ousil M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.,
RA "genome sequence of Yersinia pestis, the causative agent of plague,"
RA Nature 413:523-527(2001).
RT Neture 413:523-527(2001).
CC -1- FUNCTION: Probably involved in periplasmic detoxification of
CC copper by oxidizing Cu(II) to Cu(III) and thus preventing its uptake
CC into the cytoplasm. Possesses phenoloxidase and ferroxidase
CC activities and might be involved in the production of polyphenolic
CC compounds and the prevention of oxidative damage in the periplasm
CC (By similarity).
CC -1- COFACTOR: This protein belongs to the multicopper oxidases which
CC contain three distinct Cu centers known as type 1 or blue, type 2
CC or normal, and type 3 or coupled binuclear (By similarity).
CC -1- SUBUNIT: Monomer (Probable).
CC -1- SUBCELLULAR LOCATION: Periplasmic. It is exported via the Tat
CC pathway (By similarity).
CC -1- INDUCTION: By copper, at increased levels of cytoplasmic cuprous
CC ions (Probable).
CC -1- DOMAIN: The methionine-rich domain could provide binding sites for
CC exogenous copper ions. This methionine-rich region is probably
CC important for copper tolerance in bacteria.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A414157; CAC92639.1;
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR002355; MultiCu_oxidase2.
DR Pfam: PF00394; Cu-oxidase; 1.
DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
KW Oxidoreductase; Copper; Metal-binding; Periplasmic; Signal; Repeat;

KW Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 533
FT DOMAIN 68 164
FT DOMAIN 165 425
FT DOMAIN 426 533
FT DOMAIN 356 415
FT DOMAIN 102 102
FT METAL 104 104
FT METAL 142 142
FT METAL 144 144
FT METAL 458 458
FT METAL 461 461
FT METAL 463 463
FT METAL 514 514
FT METAL 516 516
FT METAL 520 520
FT METAL 525 525
SQ SEQUENCE 533 AA; 58328 MW; ED1570C9B9B1C35 CRC64;
Query Match 2.2%; Score 7; DB 1; Length 533;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 LAIGSQV 62
Db 324 LAIGSQV 330

RESULT 22
ERF_MOUSE STANDARD; PRT; 551 AA.
ID ERF_MOUSE
AC P70459; 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DE Blue copper oxidase cueo precursor (Copper efflux oxidase).
GN ERF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX STRAIN=129/SVJ;
RA MEDLINE=97282708; PubMed=9136988;
RA Liu D., Pavlopoulos E., Modi W., Moschonas N., Mavrothalassitis G.J.,
RT analysis of the human and mouse genes.",
RT Oncogene 14:1445-1451(1997).
RL -1- FUNCTION: POTENT TRANSCRIPTIONAL REPRESSOR THAT BINDS TO THE H1
CC ELEMENT OF THE ETS2 PROMOTER. MAY REGULATE OTHER GENES INVOLVED
CC IN CELLULAR PROLIFERATION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PHOSPHORYLATED BY MULTIPLE KINASES INCLUDING PROBABLY ERK2.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U58533; AAC09474.1;
DR EMBL: U58534; AAC09474.1; JOINED.
DR HSP: Q01543; IFLI.
DR MGD: MGI:109637; Erf.
DR InterPro: IPR000419; Ets.
DR InterPro: IPR002341; HSF_ETS.

```

DR PFam, PF00178, Ets, 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS00346; ETS_DOMAIN_3; 1.
DR TRANSCRIPTION regulation; Repressor; DNA-binding; Nuclear protein;
KW Phosphorylation.
FT DNA_BIND 27 107 ETS-DOMAIN.
FT DOMAIN 166 171 POLY-SER.
FT DOMAIN 290 293 POLY-GLY.
FT DOMAIN 362 373 POLY-SER.
FT DOMAIN 420 425 POLY-PRO.
FT MOD_RES 529 529 PHOSPHORYLATION (BY MAPK1) (BY
SIMILARITY).
SQ SEQUENCE 551 AA; 59050 MW; 5AC1B72FB2743FE5 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 SSGSGGA 57
DB 406 SSGSGGA 412

RESULT 23
FIXN AZOCA STANDARD; PRT; 551 AA.
AC P98056;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome c oxidase polypeptide I homolog (EC 1.9.3.1).
GN FIXN.
OS Azorhizobium caulinodans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hyphomicrobium group; Azorhizobium.
OX NCBI_TaxId=7;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORS571;
RA MEDLINE=94109675; PubMed=8282187;
RA Mandon K., Kaminski P.A., Mougel C., Desnoues N., Eimerich C.;
RT "Role of the fixXII region of Azorhizobium caulinodans in free-living
and symbiotic nitrogen fixation.";
RL FEMS Microbiol. Lett. 114:185-189(1993).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C OR A QUINOL ARE TRANSFERRED TO THE BINETALLIC CENTER
FORMED BY A HIGH-SPIN HEME AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome
c + 2 H(2)O.
CC -1- PATHWAY: Respiratory chain; terminal step.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 13 OR 14
POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; X74410; CAA52429.1; -
CC InterPro; IPR0000883; COX1.
CC InterPro; IPR004677; CCON.
CC PFam; PF00115; COX1; 1.
CC TIGRfam; TIGR00780; CCON; 1.

```

```

DR PROSITE; PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.
FT METAL 132 132 IRON (LOW-SPIN HEME) (BY SIMILARITY).
FT METAL 281 281 COPPER B (BY SIMILARITY).
FT METAL 331 331 COPPER B (BY SIMILARITY).
FT METAL 332 332 COPPER B (BY SIMILARITY).
FT METAL 419 419 IRON (HIGH-SPIN HEME) (BY SIMILARITY).
FT METAL 421 421 IRON (LOW-SPIN HEME) (BY SIMILARITY).
SQ SEQUENCE 551 AA; 61843 MW; 0466F71CD04BFC4D CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 OPSGAL 303
DB 541 OPSGAL 547

RESULT 24
UL25_HSV1 STANDARD; PRT; 580 AA.
AC P10209;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Varion protein UL25.
GN UL25.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxId=10299;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
herpes simplex virus type 1.";
RL J. Gen. Virol. 69:1531-1574(1988).
CC -1- FUNCTION: VIRION PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL25,
EHV-1 36, EBV BFRF1, HCMV UL77, ILTV ORF2, AND VZV 34.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; D10879; BAA01671.1; -
CC EMBL; X14112; CAA32317.1; -
CC PIR; G30084; WMBEWS.
CC InterPro; IPR002493; UL25.
CC PFam; PF01499; UL25; 1.
SQ SEQUENCE 580 AA; 62669 MW; 3F8F0B7C122B2536 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 ATCGSGV 197
DB 167 ATCGSGV 173

RESULT 25
SPPA_ECOLI STANDARD; PRT; 618 AA.
ID SPPE_ECOLI
AC P08395; P77752; Q46724; Q46725; Q46726; Q57183;
DT 01-AUG-1988 (Rel. 08, Created)

```

DT DT 01-AUG-1988 (Rel. 08, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Proclease IV (EC 3.4.21.-) (Endopeptidase IV) (Signal peptide
peptide).
GN SPFA OR B1766.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86250892; PubMed=3522590;
RA Ichihara S., Suzuki T., Suzuki M., Mizushima S.,
RT "Molecular cloning and sequencing of the spn gene and
characterization of the encoded protease IV, a signal peptide
peptidase, of Escherichia coli,"
RL J. Biol. Chem. 261:9405-9411(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655,
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RL "The complete genome sequence of Escherichia coli K-12";
RT Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Mihi T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Naishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seti Y., Sivastandarum S., Tagami H., Takeeda J.,
RA Takekoshi K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.",
RL DNA Res. 3:363-377(1996).
RN [4]
RP SEQUENCE OF 110-433 FROM N.A.
RC STRAIN=various ECOL strains,
RX MEDLINE=95064015; PubMed=7937328;
RA Gutman D.S., Dykhuisen D.E.;
RT "Clonal divergence in Escherichia coli as a result of recombination,
not mutation.";
RL Science 266:1380-1383(1994).
CC -1- FUNCTION: DIGESTION OF THE CLEAVED SIGNAL PEPTIDES. THIS ACTIVITY
IS NECESSARY TO MAINTAIN PROPER SECRETION OF MATURE PROTEINS
ACROSS THE MEMBRANE.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(Probable).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S49.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

DR EMBL: AL03359; AAA24648.1; -
DR EMBL: AE000271; AAC74836.1; -
DR EMBL: D90820; PA01557.1; -
DR EMBL: U13772; AA057008.1; -
DR EMBL: U13773; AA057009.1; -
DR EMBL: U13774; AA057010.1; -
DR EMBL: U13775; AA057011.1; -
DR EMBL: U13776; AA057012.1; -
DR EMBL: U13777; AA057013.1; -

Query Match 378 378 6723 MW; EMB51C3E2946AC3 CRC64;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 434 IFGVIT 440

21 IFGVIT 27
|||||

RESULT 26
PART1 YEAST
ID PART1 YEAST STANDARD; PRT; 623 AA.
AC P36223;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable long-chain fatty acid transport protein.
GN PART1 OR YBR041W OR YBR0411.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=5288C;
RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Uirestarazu A.,
RL Vassers S.;
RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RP STRAIN=M303A.
RX MEDLINE=97236810; PubMed=9079682;
RA Ferguson N.J., Dinuso C.C., Elberger A., Knudsen J., Black P.N.;
RT "Disruption of the Saccharomyces cerevisiae homologue to the murine
RT fatty acids transpore protein impairs uptake and growth on long-chain
RT fatty acids." 272.8531-8538 (1997).
RL J. Biol. Chem. 272.8531-8538 (1997).
CC -1- FUNCTION: MAY BE INVOLVED IN LONG-CHAIN FATTY ACIDS UPTAKE, AND
CC THUS MAY PLAY A PIVOTAL ROLE IN REGULATING THEIR ACCESSIBILITY
CC PRIOR TO METABOLIC UTILIZATION. MAY PLAY AN IMPORTANT ROLE IN
CC UPTAKE OF THESE HYDROPHOBIC COMPOUNDS UNDER CONDITIONS WHERE FATTY
CC ACID SYNTHESIS IS COMPROMISED.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z35910; CAA84983.1; -.
CC PIR; S45899; S45899.
CC SGD; S0000245; PAT1.
CC InterPro; IPR000873; AMP-bind.
CC Pfam; PF00501; AMP-binding; 1.
CC PROSITE; PS00455; AMP_BINDING; 1.
CC KW Lipid transport; Transmembrane.
CC FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 54 71 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 623 AA; 71697 MW; 0AE02700B60C8CFC CRC64;

Query Match 2.2%; Score 7; DB 1; Length 623;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPPAT 192
Db 398 TEAPPAT 404
|||||
DCMA MOOTH STANDARD; PRT; 729 AA.
ID DCMA MOOTH STANDARD; PRT; 729 AA.
AC P27988;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbon monoxide dehydrogenase alpha subunit (EC 1.2.99.2) (COOH).
OS Moorella thermoacetica (Clostridium thermoacetum).
OC Bacteria; Firmicutes; Clostridia; Thermotoga bacteriales;
OC Thermotogaceae; Thermotogales; Moorella group; Moorella.
OX NCBI_TaxID=1525;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92084676; PubMed=1748656;
RA Morton T.A., Runquist J.A., Ragsdale S.W., Shanmuganandaram T.,
RA Wood H.G., Ljungdahl L.G.;
RT "The primary structure of the subunits of carbon monoxide
RT dehydrogenase/acyl-CoA synthase from Clostridium thermoacetum";
RT J. Biol. Chem. 266:23824-23828(1991).
RL -1- FUNCTION: CATALYZES THE INTERCONVERSION OF CO AND CO2 AND THE
CC SYNTHESIS OF ACETYL-COENZYME A FROM THE METHYLATED CORRIDOID/IRON
CC SULFUR PROTEIN, CO, AND COENZYME A.
CC -1- CATALYTIC ACTIVITY: CO + H(2)O + acceptor = CO(2) + reduced
CC acceptor.
CC -1- COFACTOR: NICKEL.
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; M62727; AAA23229.1; -.
CC PIR; B41670; B41670.
CC InterPro; IPR004461; CdhC.
CC Pfam; PF03598; CdhC; 1.
CC TIGRfam; TIGR00316; CdhC; 1.
CC KW Oxidoreductase; Nickel.
SQ SEQUENCE 729 AA; 81725 MW; 619BB19D959F5A72 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 729;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 PVPYRFG 215
Db 139 PVPYRFG 145
|||||
YKGT CAEEL STANDARD; PRT; 781 AA.
ID YKGT CAEEL STANDARD; PRT; 781 AA.
AC P46557;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 86.0 kDa protein B0285.7 in chromosome III precursor.
GN B0285.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Sulston J.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z34533; CAA84298.1; -.
CC WormPep; B0285.7; CE00646.
CC InterPro; IPR001930; Ala_peptase.
CC Pfam; PF01433; Peptidase_M1; 1.
CC KW Hypothetical protein; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 781 HYPOTHETICAL PROTEIN B0285.7.
FT DOMAIN 70 73 POLY-THR.
FT DOMAIN 81 84 POLY-THR.
FT DOMAIN 89 93 POLY-THR.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 781 AA; 85984 MW; A9D5BCD8F567B8B0 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 781;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 SISSGVN 132
Db 257 SISSGVN 263
|||||
CPN DROME
RESULT 29

ID CPN DROME STANDARD; PRT; 865 AA.
 AC 002510;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Calphotin.
 GN CPN OR CAP.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Cancon-S;
 RC MEDLINE=93165729; PubMed=8094559;
 RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
 RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Cancon-S;
 RC MEDLINE=93165730; PubMed=8434015;
 RA Ballinger D.G., Xue N., Harshman K.D.;
 RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
 RL calcium and contains a leucine zipper";
 CC Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
 CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
 CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
 CC OF CA-2 PER MOL OF PROTEIN.
 CC -1- SUBUNIT: HOMODIMER (PROBABLY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
 CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
 CC COMPOUND EYES AND OCELLI.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
 CC DEVELOPMENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L02111; AAA28405.1; -;
 DR EMBL; L05080; AAA28420.1; -;
 DR PIR; A47282; A47282.
 DR FlyBase; FBgn010218; Cpn.
 KM Calcium-binding.
 FT CONFLICT 36 A -> AVAPAVVA (IN REF. 2).
 FT CONFLICT 43 I -> T (IN REF. 2).
 FT CONFLICT 64 I -> V (IN REF. 2).
 FT CONFLICT 76 T -> A (IN REF. 2).
 FT CONFLICT 100 P -> pp (IN REF. 2).
 FT CONFLICT 126 VQ -> AP (IN REF. 2).
 FT CONFLICT 154 I -> V (IN REF. 2).
 FT CONFLICT 160 S -> T (IN REF. 2).
 FT CONFLICT 534 S -> E (IN REF. 2).
 FT CONFLICT 534 I -> T (IN REF. 2).
 FT CONFLICT 699 V -> L (IN REF. 2).
 FT CONFLICT 703 D -> E (IN REF. 2).
 FT CONFLICT 721 D -> E (IN REF. 2).
 SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFC CRC64;
 Query Match 2.28; Score 7; DB 1; Length 865;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 30
 HIRA DROME STANDARD; PRT; 1047 AA.
 ID HIRA DROME
 AC 017468; OSW303; 046105; 077144;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE HIRA protein homolog (HIRA).
 GN HIRA OR PNH OR CG12153
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RC MEDLINE=98278847; PubMed=9611274;
 RA Kirov N., Shitilbans A., Rushlow C.;
 RT "Isolation and characterization of a new gene encoding a member of the
 RL HIRA family of proteins from Drosophila melanogaster.";
 RN Gene 212:323-332(1998).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Embryo;
 RC MEDLINE=98380288; PubMed=9712723;
 RA Llevadot R., Marques G., Pritchard M., Estivill X., Ferrus A.,
 RA Scambler P.;
 RT "Cloning, chromosome mapping and expression analysis of the HIRA gene
 RL from Drosophila melanogaster.";
 RN Biochem. Biophys. Res. Commun. 249:486-491(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RC STRAIN=Berkley;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton R.G., Wortman J.R., Tangell M.D., Zheng X., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazey R.G., Champ M., Felfler B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.I.G.,
 RA April J.F., Agapayni A., An H.-U., Andrews-Ramnoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harrit N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markov G., Mileshina N.V., Mobarry C., Morris J., Mosnere D.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon S., Nollard J., Puri V., Reese M.G.,
 RA Palazzolo M., Peltman G.S., Pan S., Sanders R.D., Schaefer F., Smith H.,
 RA Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,
 RA Shue B.C., Siden-Kiamos I., Stimpson W., Strong R., Sun E.,
 RA Spier E., Spradling A.C., Stapleton M., Turner R., Venter E., Wang X.H., Wang X.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang X.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng J.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";

[illegible]

Query Match	Score 7	DB 1	Length 1256
Best Local Similarity	100.0%	Pred No. 98	
CC	01-OCT-1996 (Ref. 34, Created)		
DT	01-OCT-1996 (Ref. 34, Last sequence update)		
DT	15-JUN-2002 (Ref. 41, Last annotation update)		
DE	Bifunctional autolysin precursor [Includes: N-acetylmuramoyl-L-alanine		
DE	amidase (EC 3.5.1.28) ; Mannoysl-glycoprotein endo-beta-N-		
DE	acetylglucosamidase (EC 3.2.1.96)] .		
GN	ATL.		
OS	Staphylococcus aureus .		
CC	Bacteria; Firmicutes; Bacillales; Staphylococcus .		
OX	NCBI_TaxID=1280;		
RN	[1]		
RP	SEQUENCE FROM N.A. , AND SEQUENCE OF 205-214 AND 776-792 .		
RC	STRAIN=RH450;		
EX	MEDLINE=95116542; PubMed=7816834 .		
RA	Onidra T. , Sugai M. , Komatsuzaawa H. , Hong Y.-M. , Sugitaka H. ,		
RA	Tomasz A. ;		
RT	"A staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-		
RT	alanine amide domain and an endo-beta-N-acetylglucosaminidase		
RT	domain: cloning, sequence analysis, and characterization." ;		
RL	Proc. Natl. Acad. Sci. U.S.A. 92:285-289 (1995) .		
RN	[2]		
RP	SEQUENCE FROM N.A. .		
RC	STRAIN=NCCTC 8325-4;		
RA	Poster S.J. ;		
RL	Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases .		
CC	-1- FUNCTION: ENDOLYOLYSIS OF THE DI-N-ACETYLCHITOSYL UNIT IN		
CC	HIGH-MANNOSE GLYCOPETIDES AND GLYCOPROTEINS CONTAINING THE		
CC	-1-(MAN) 5 (GLCNAC) 2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE		
CC	RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE		
CC	OLIGOSACCHARIDE IS RELEASED INTACT.		
CC	-1- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl		
CC	residues and L-amino acid residues in certain bacterial cell-wall		
CC	glycopeptides.		
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl		
CC	unit in high-mannose glycopeptides and glycoproteins containing		
CC	the -[Man(GlcNAc)2]asn-structure. One N-acetyl-D-glucosamine		
CC	residue remains attached to the protein; the rest of the		
CC	oligosaccharide is released intact.		
CC	-1- SUBCELLULAR LOCATION: Secreted.		
CC	-1- SUBCELLULAR LOCATION: Secreted.		
CC	-1- PM: UNDECODED PROTEOLYTIC PROCESSING TO GENERATE THE TWO		
CC	EXTRACELLULAR LYTIC ENZYMES.		
CC	-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-		
CC	ACETYLURAMYOYL-L-ALANINE AMIDASE FAMILY 2 .		
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF		
CC	GLYCOSYL HYDROLASES.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.1sb-sib.ch/announce/		
CC	or send an email to license@1sb-sib.ch).		
CC	-----		
DR	EMBL; D17366; BAA04185.1; -		
DR	EMBL; L41499; AAA99982.1; -		
DR	InterPro; IPR002502; Amidase_2 .		
DR	InterPro; IPR002901; Amidase_4 .		
DR	Pfam; PF01510; Amidase_2; 1 .		
DR	Pfam; PF01832; Amidase_4; 1 .		
DR	SMART; SM00047; LY22; 1 .		
KW	Cell wall; Hydrolyase; Signal; multifunctional enzyme; Repeat.		
FT	SIGNAL 1 29		POTENTIAL.
FT	CHAIN 30 1256		BIFUNCTIONAL AUTOLYSIN.
FT	DOMAIN 199 775		N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT	DOMAIN 776 1256		ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
FT	REPEAT 425 589		1 .
FT	REPEAT 596 758		2 .
FT	REPEAT 770 932		3 .
SQ	SEQUENCE 1256 AA; 137384 MW; 2BB76CA292PDD20 CRC64;		

RA Gwin M., Nelson W., DeBoy R., Kolney J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genomic sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002338; AAF39510.1; -.
CC TIGR; TC0694; -.
DR InterPro; IPR003368; Chlamydia_PMP.
DR InterPro; IPR003357; OMP.
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 3.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 1672 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
SQ SEQUENCE 1672 AA; 176295 MW; 4ABF190DA4DF8B06 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 1672;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PTPPTPV 174
DB 404 PTPPTPV 410

RESULT 34
YS89 CAEEL STANDARD; PRT; 3178 AA.
AC 009624; 009625; 0969D4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK945.9 in chromosome II.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RA Wilkinson-Spratt J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48544; CAB70192.1; -.
CC EMBL; Z48582; CAB70192.1; JOINED.
DR EMBL; Z48582; CAB70201.1; -.

DR EMBL; Z48544; CAB70201.1; JOINED.
DR WormRep; ZK945.9; CE25697.
DR InterPro; IPR002111; Cat channel TrpL.
DR InterPro; IPR001024; Lipoygenase LH2.
DR InterPro; IPR000636; M-channel nlg.
DR InterPro; IPR000203; PKD_cyb_rich.
DR Pfam; PF00520; Ion_trans; 1.
DR Pfam; PF01477; PLAT; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00308; LH2; 1.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 266 1196 SER/THR-RICH.
FT DOMAIN 1105 1241 GLY/SER-RICH.
FT DOMAIN 2071 2120 GPS.
FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 51 73 POTENTIAL.
FT TRANSMEM 2139 2161 POTENTIAL.
FT TRANSMEM 2348 2367 POTENTIAL.
FT TRANSMEM 2390 2412 POTENTIAL.
FT TRANSMEM 2451 2468 POTENTIAL.
FT TRANSMEM 2483 2505 POTENTIAL.
FT TRANSMEM 2567 2589 POTENTIAL.
FT TRANSMEM 2586 2858 POTENTIAL.
FT TRANSMEM 2939 2961 POTENTIAL.
FT TRANSMEM 2976 2998 POTENTIAL.
FT TRANSMEM 3038 3060 POTENTIAL.
SQ SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;

Query Match 2.2%; Score 7; DB 1; Length 3178;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 1206 STSGSGS 1212.

RESULT 35
DBH1 BIFLO STANDARD; PRT; 27 AA.
ID DBH1_BIFLO
AC P17615;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNA-binding protein HBI (Fragment).
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=1679;
RN [1]
RP SEQUENCE.
RC STRAIN=ATCC 15703;
RX MEDLINE=90344917; PubMed=2116910;
RA Goshima N., Kano Y., Imamoto F.;
RT "Characterization of HU-like protein from Bifidobacterium longum.";
RL Biochimie 72:207-212(1990).
CC -1- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC PIR; A43768; A43768.
DR InterPro; IPR00119; Bac_DNABind.
DR Pfam; PF00216; Bac DNA binding; 1.
DR PROSITE; PS00045; HISTONE LIKE; PARTIAL.
KW DNA-binding; DNA condensation.
FT NON TER 27
SQ SEQUENCE 27 AA; 2863 MW; 6A201A5965A1BD64 CRC64;

Query Match 1.9%; Score 6; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 33;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 VSKIAQ 86
Db 8 VSKIAQ 13

RESULT 36
PRTE ONCMY STANDARD; PRT: 32 AA.

AC P08145;
DT 01-FEB-1998 (Rel. 08, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Proteasome P14.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP MEDLINE=83272939; PubMed=6308564;
RA Alken J.M., McKenzie D., Zhao H.-Z., States J.C., Dixon G.H.;
RT "Sequence homologies in the proteasome gene family of rainbow trout."
RL Nucleic Acids Res. 11:4907-4922(1983).
CC -1- FUNCTION: PROTEASOMES SUBSTITUTE FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: X01595; CAA25748.1; -.
DR PIR: A21211; A21211.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Tests; DNA condensation; Nuclear protein.
FT INIT MET 0
SQ SEQUENCE 32 AA; 4260 MW; 5D388D9D90B796E9 CRC64;

Query Match 1.9%; Score 6; DB 1; Length 32;
Best local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 144 SSRPPV 149
Db 6 SSRPPV 11

RESULT 37
YSXC_SULAC STANDARD; PRT: 40 AA.

AC P39477;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 4.5 kDa protein in SOXC 3' region.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RA MEDLINE=92192013; PubMed=1372250;
RA Luebben M., Kolmerer B., Saraste M.;

RT "An archaeobacterial terminal oxidase combines core structures of two
RT mitochondrial respiratory complexes.";
RL EMBL J. 11:805-812(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: X62643; CAA44512.1; -.
DR PIR: S21044; S21044.
KW Hypothetical protein.
SQ SEQUENCE 40 AA; 4502 MW; 5EEA6A276C0B8A1D CRC64;

Qy 18 LGLIFG 23
Db 9 LGLIFG 14

RESULT 38
IVBC_NAJNA STANDARD; PRT: 57 AA.

AC P19859;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Venom chymotrypsin inhibitor.
OS Naja naja (Indian cobra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosaurs; Squamata; Scleroglossae; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=35670;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=91085579; PubMed=2262001;
RA Snafgat J., Zaidi Z.H., Joernvall H.;
RT "Purification and characterization of a chymotrypsin Kunitz inhibitor
RT type of polypeptide from the venom of cobra (Naja naja naja)."
RL FEBS Lett. 275:6-8(1990).
CC -1- FUNCTION: THIS PROTEIN INHIBITS CHYMOTRYPSIN.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC -----
CC PIR: S12957; S12957.
DR HSBP; P00981; IDTK.
DR InterPro; IPR002223; Kunitz BPTI.
DR Pfam; PF00014; Kunitz BPTI; 1.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00131; Ku; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_2; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_1.
KW Venom; Serine protease inhibitor.
FT DISULFID 5
FT DISULFID 35
FT DISULFID 14
FT DISULFID 30
FT ACT_SITE 15
SQ SEQUENCE 57 AA; 6508 MW; 4BFB26A010C1737A CRC64;

Query Match 1.9%; Score 6; DB 1; Length 57;
Best local Similarity 100.0%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 GFVSSY 271
Db 16 GFVSSY 21

```

RESULT 39
YL3_ADE41
ID YL3_ADE41 STANDARD; PRT; 60 AA.
AC P23650;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 6.9 kDa protein in 100 kDa protein region.
OS Human adenovirus type 41.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OK NCBI_TaxID=10524;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok;
RX MEDLINE=90272433; PubMed=2349115;
RA Slemenda S.B., Pieniazek N.U., Velarde J. Jr., Pieniazek D.,
RA Luftig R.B.;
RT "Nucleotide sequence of the region coding for 100K and 33K proteins
RT of human enteric adenovirus type 41 (Tak).";
RL Nucleic Acids Res. 18:3069-3069(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52532; CAA36763.1; -
DR PIR; S10210; S10210.
KM Hypothetical protein.
SQ SEQUENCE 60 AA; 6897 MW; 9AC2E2D25A196138 CRC64;

Query Match
Best Local Similarity 1.9%; Score 6; DB 1; Length 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AVQSSR 146
DB 21 AVQSSR 26

RESULT 40
EC_MAIZE
ID EC_MAIZE STANDARD; PRT; 76 AA.
AC P43401;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE EC protein homolog (Zinc-metallothionein class II).
OS Zea mays (maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OK NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA White C.N., Rivin C.J.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS 5 MOLECULES OF ZINC. MAY HAVE A ROLE IN ZN(2+)
CC HOMEOSTASIS DURING EMBRYOGENESIS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; Z34469; CAA84233.1; -
DR EMBL; U10696; AAA19405.1; -
DR MaizeDB; 78896; -
DR InterPro; IPR000316; Metallothio15.
DR Pfam; PF02068; Metallothio_PEC; 1.
DR PRINTS; PR00877; MTPANTPEC.
KM Metal-binding; Metal-chiolate cluster; Zinc.
INIT MET 0
SQ SEQUENCE 76 AA; 7574 MW; C9803CBF7F9B1D34 CRC64;

Query Match
Best Local Similarity 1.9%; Score 6; DB 1; Length 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 TSGSGG 55
DB 21 TSGSGG 26

Search completed: July 8, 2003, 11:19:16
Job time : 43 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:14:23 ; Search time 80 Seconds
(without alignments)
829.340 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 322
Sequence: 1 MVTIAINSQNKPIKRLGL.....LFEFRISNGVYDPTLVK 322

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.8	968	5 Q9VCB2	Q9VCB2 drosophila
2	8	2.5	91	10 Q8S0D0	Q8S0D0 oryza sativ
3	8	2.5	120	10 Q42049	Q42049 arabidopsis
4	8	2.5	164	16 Q34816	Q34816 bacillus su
5	8	2.5	227	16 Q9K644	Q9K644 bacillus ha
6	8	2.5	244	3 Q05905	Q05905 saccharomyc
7	8	2.5	306	10 Q9SET1	Q9SET1 arabidopsis
8	8	2.5	334	10 Q9LIE9	Q9LIE9 arabidopsis
9	8	2.5	376	10 Q39353	Q39353 brassica na
10	7	2.2	88	4 Q9LUC5	Q9LUC5 homo sapien
11	7	2.2	101	16 Q9CDX3	Q9CDX3 lactococcus
12	7	2.2	111	10 Q8S142	Q8S142 oryza sativ
13	7	2.2	122	10 Q8RUE2	Q8RUE2 oryza sativ
14	7	2.2	133	10 Q9ZWI6	Q9ZWI6 zea mays (m
15	7	2.2	133	10 Q9ST25	Q9ST25 zea mays (m
16	7	2.2	136	10 Q9FWP4	Q9FWP4 oryza sativ

17	7	2.2	141	5 P91181	P91181 caenorhabdi
18	7	2.2	142	10 Q9FWP0	Q9FWP0 oryza sativ
19	7	2.2	165	17 Q8THE9	Q8THE9 methan sarc
20	7	2.2	176	16 Q926C3	Q926C3 rhizobium m
21	7	2.2	177	5 Q9XW11	Q9XW11 caenorhabdi
22	7	2.2	179	16 Q99YU1	Q99YU1 streptococc
23	7	2.2	180	6 Q95KR4	Q95KR4 sus scrofa
24	7	2.2	189	16 Q8ZFP8	Q8ZFP8 yersinia pe
25	7	2.2	195	5 Q9VJ78	Q9VJ78 drosophila
26	7	2.2	205	16 Q8XXR6	Q8XXR6 ralsstonia s
27	7	2.2	208	10 Q81130	Q81130 arabidopsis
28	7	2.2	208	10 Q9SFD8	Q9SFD8 arabidopsis
29	7	2.2	211	16 Q9S2U3	Q9S2U3 streptococc
30	7	2.2	212	10 Q9FYQ4	Q9FYQ4 oryza sativ
31	7	2.2	213	5 Q9VY29	Q9VY29 drosophila
32	7	2.2	216	8 Q8SN98	Q8SN98 euglena gra
33	7	2.2	225	17 Q92808	Q92808 archaeoglob
34	7	2.2	226	16 Q92KU4	Q92KU4 rhizobium m
35	7	2.2	228	4 Q9H5P8	Q9H5P8 homo sapien
36	7	2.2	241	16 Q9KLV7	Q9KLV7 vibrio chol
37	7	2.2	241	17 Q8RUT2	Q8RUT2 methanobact
38	7	2.2	249	17 Q9HNY3	Q9HNY3 halobacteri
39	7	2.2	259	16 Q51373	Q51373 borrelia bu
40	7	2.2	269	5 Q94238	Q94238 caenorhabdi
41	7	2.2	276	10 Q9SPD9	Q9SPD9 zea mays (m
42	7	2.2	276	10 Q9SPD8	Q9SPD8 zea mays (m
43	7	2.2	281	16 Q9EMD9	Q9EMD9 streptococc
44	7	2.2	285	17 Q2S731	Q2S731 methanobact
45	7	2.2	288	2 Q52246	Q52246 enterococcu
46	7	2.2	291	16 Q985F0	Q985F0 rhizobium 1
47	7	2.2	302	12 Q91HNS	Q91HNS human polio
48	7	2.2	303	10 Q41814	Q41814 zea mays (m
49	7	2.2	309	16 Q98L55	Q98L55 rhizobium 1
50	7	2.2	309	16 Q8Y158	Q8Y158 brucella me
51	7	2.2	310	2 Q51853	Q51853 pseudomonas
52	7	2.2	312	2 Q9K4R9	Q9K4R9 pseudomonas
53	7	2.2	312	11 Q8VGC0	Q8VGC0 mus musculu
54	7	2.2	326	4 Q62449	Q62449 caenorhabdi
55	7	2.2	328	5 Q96E29	Q96E29 homo sapien
56	7	2.2	328	4 Q8WYX0	Q8WYX0 homo sapien
57	7	2.2	328	10 Q42366	Q42366 zea mays (m
58	7	2.2	329	9 Q74489	Q74489 schizosacch
59	7	2.2	334	11 Q9D205	Q9D205 mus musculu
60	7	2.2	339	5 Q18567	Q18567 caenorhabdi
61	7	2.2	344	2 Q82858	Q82858 acetobacter
62	7	2.2	350	10 Q41719	Q41719 zea diptope
63	7	2.2	351	16 Q8ZDT2	Q8ZDT2 yersinia pe
64	7	2.2	355	16 Q9ZM09	Q9ZM09 helicobacte
65	7	2.2	361	5 Q9N1P2	Q9N1P2 drosophila
66	7	2.2	368	5 Q9XUK0	Q9XUK0 caenorhabdi
67	7	2.2	369	10 Q40692	Q40692 oryza sativ
68	7	2.2	374	5 Q9N1H3	Q9N1H3 drosophila
69	7	2.2	377	16 Q97DU6	Q97DU6 clostridium
70	7	2.2	380	16 Q98180	Q98180 rhizobium 1
71	7	2.2	381	5 Q94399	Q94399 caenorhabdi
72	7	2.2	387	5 Q16316	Q16316 caenorhabdi
73	7	2.2	387	16 Q8YFP9	Q8YFP9 anabaena sp
74	7	2.2	388	17 Q97B14	Q97B14 thermoplasma
75	7	2.2	389	2 Q9X9P9	Q9X9P9 streptococc
76	7	2.2	390	5 Q9GRY4	Q9GRY4 caenorhabdi
77	7	2.2	391	16 Q9PD30	Q9PD30 xylella fas
78	7	2.2	392	5 Q9W5S7	Q9W5S7 drosophila
79	7	2.2	393	5 Q9N2Z3	Q9N2Z3 caenorhabdi
80	7	2.2	394	5 Q9GNA3	Q9GNA3 drosophila
81	7	2.2	395	5 Q9GNJ2	Q9GNJ2 drosophila
82	7	2.2	395	5 Q9V181	Q9V181 drosophila
83	7	2.2	395	5 Q9U4F5	Q9U4F5 drosophila
84	7	2.2	395	5 Q9U6N4	Q9U6N4 drosophila
85	7	2.2	396	5 Q44435	Q44435 drosophila
86	7	2.2	396	5 Q9W2S5	Q9W2S5 drosophila
87	7	2.2	400	2 Q938U4	Q938U4 streptococc
88	7	2.2	400	16 Q9A0M2	Q9A0M2 streptococc
89	7	2.2	409	5 Q9GND0	Q9GND0 drosophila

966	6	1.9	281	10	094Bt6	094Bt6 ceratostigm
967	6	1.9	281	10	022842	022842 arabidopsis
968	6	1.9	281	16	082AK7	082AK7 yersinia pe
969	6	1.9	282	10	09GCS2	09GCS2 oryza sativ
970	6	1.9	283	16	09A4H3	09A4H3 caulobacter
971	6	1.9	284	10	047809	047809 enterococcu
972	6	1.9	284	10	09A0G4	09A0G4 arabidopsis
973	6	1.9	284	16	08ZDU0	08ZDU0 yersinia pe
974	6	1.9	284	16	08ZDU0	08ZDU0 yersinia pe
975	6	1.9	284	17	0972X1	0972X1 sulfolobus
976	6	1.9	285	17	097750	097750 uncultured
977	6	1.9	286	10	09BMC0	09BMC0 venturia ca
978	6	1.9	286	11	09DAC6	09DAC6 mus musculu
979	6	1.9	286	16	06B342	06B342 clover yell
980	6	1.9	286	16	098DW5	098DW5 rhizobium 1
981	6	1.9	286	16	09ZC23	09ZC23 yersinia pe
982	6	1.9	286	16	09K478	09K478 streptomyce
983	6	1.9	287	5	08SRA0	08SRA0 encephalito
984	6	1.9	287	10	08JY14	08JY14 arabidopsis
985	6	1.9	287	11	08VH31	08VH31 mus musculu
986	6	1.9	288	12	06V341	06V341 clover yell
987	6	1.9	288	12	09DMD6	09DMD6 rat cytochrome
988	6	1.9	288	16	091480	091480 pseudomonas
989	6	1.9	288	17	09HRF8	09HRF8 halobacteri
990	6	1.9	290	2	093CK9	093CK9 mycoplasma
991	6	1.9	290	16	08XW8	08XW8 clostridium
992	6	1.9	290	16	08UJK4	08UJK4 agrobacteri
993	6	1.9	291	4	096AD3	096AD3 homo sapien
994	6	1.9	291	10	024273	024273 polygonatum
995	6	1.9	291	10	08W318	08W318 vitis labru
996	6	1.9	291	16	09POA2	09POA2 ureaplasma
997	6	1.9	292	11	08VDB7	08VDB7 mus musculu
998	6	1.9	292	13	09DPC2	09DPC2 xenopus lae
999	6	1.9	292	16	092P88	092P88 listeria mo
1000	6	1.9	292	17	058756	058756 methanococc

ALIGNMENTS

RESULT 1
ID 09VCB2; PRELIMINARY; PRT; 968 AA.
AC 09VCB2; 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE CG5669 protein (L04007p).
GN CG5669.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Gallie R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miller G.L.,
RA Abiri J.F., Agbayan A., An H.-U., Andrews-Piankovich C., Baldwin D.,
RA Baller R.M., Basu A., Bendale U., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegawa C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.L.,
RA Merklion G., Mikhlin N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy S., Murphy I., Murry D.M., Nelson D.L.,
RA Nelson S.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo R., Feltman G.S., Fan S., Foland J., Fui V., Reese M.G.,
RA Reineck K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guatin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacle J., Paragas V., Park S.,
RA Patel S., Phoumenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celisner S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: ARO03746; AAF56261.1; -
DR EMBL: AYO89533; AAL90271.1; -
DR HSSP: P08047; 1SP2.
DR FlyBase: FBgn0039169; CG5669.
DR InterPro: IPR000823; Znf_C2H2.
DR Pfam: PF00086; Znf_C2H2_3.
DR ProDom: PD000003; Znf_C2H2; 2.
DR SMART: SM00355; Znf_C2H2; 3.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 968 AA; 102020 MW; 834CB19340C302CB CRC64;
Query Match 2.8%; Score 9; DB 5; Length 968;
Best Local Similarity 100.0%; Pred. No. 2.9; Mismatches 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Gaps 0;
Oy 217 ATVAGSTVT 225
Db 877 ATVAGSTVT 885
RESULT 2
ID 08SOD0; PRELIMINARY; PRT; 91 AA.
AC 08SOD0; 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE P050610.26 protein.
GN Oryza sativa (Japanese cultivated group).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nippondare(GA3) genomic DNA, chromosome 1, PAC
RT clone:PO506A10.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003418; BAB86093.1; -
SQ SEQUENCE 91 AA; 9299 MW; 06773A6E70E18FE8 CRC64;

Query Match
Best Local Similarity 2.5%; Score 8; DB 10; Length 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 192 TGSSGVMQ 199
29 TGSSGVMQ 36

RESULT 3
042049 PRELIMINARY; PRT; 120 AA.
AC 042049;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE ADP1 protein (Fragment).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEEDLING;
RA Desprez T., Amselem J., Chiapello H., Caboche M., Hofte H.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z25716; CAAB1017.1; -
DR HSRP; P24337; IHP.
DR InterPro; IPR002365; P-rich_extensin.
DR InterPro; IPR001768; TRY/amy1_inhbr.
DR Pfam; PF00234; TRY_alpha_amy1; 1.
DR PRINTS; PR01217; PRICHTEXTENGN.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 120 AA; 12284 MW; BE67AECB042CB451 CRC64;

Query Match
Best Local Similarity 2.5%; Score 8; DB 10; Length 120;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 PTPPPVV 175
DB 18 PTPPPVV 25

RESULT 4
034816 PRELIMINARY; PRT; 164 AA.
ID 034816;
AC 034816;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE YKUD protein.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Scanlan E., Devine K.M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;
RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruechi C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dueterhoff A., Ehrlich S.D., Emerson P.T.,
RA Dentan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seior S.J., Setor P., Shin B.S., Soldo B.,
RA Sorokin A., Taccini B., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipac A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunze F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ222587; CA10867.1; -
DR EMBL; Z99111; CAB13277.1; -
DR InterPro; IPR002482; Lyam.
DR Pfam; PF01476; Lyam; 1.
DR SMART; SM00257; Lyam; 1.
KW Complete proteome.
SQ SEQUENCE 164 AA; 17643 MW; 6E13A750890E7F4C CRC64;

Query Match
Best Local Similarity 2.5%; Score 8; DB 16; Length 164;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 YOVKQGD 80
DB 4 YOVKQGD 11

RESULT 5
09K644 PRELIMINARY; PRT; 227 AA.
ID 09K644;
AC 09K644;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein BH3886.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*
 RT *halodurans* and genomic sequence comparison with *Bacillus subtilis*."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AF001520; BAB07607.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 227 AA; 25714 MW; EEBC5FACDC3D6F3 CRC64;

Query Match 2.5%; Score 8; DB 16; Length 227;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 ATNPVVR 213
 DB 159 ATNPVVR 166

RESULT 6
 ID 005905 PRELIMINARY; PRT; 244 AA.
 AC 005905;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Chromosome XII COSMID 8003.
 CN YLR301W OR L8003.2.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97313257; Pubmed=9169871;
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
 RA Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
 RA Emtian K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Heuss-Nelitzel D., Hilbert H., Hliger F., Kleene K., Kotter P.,
 RA Louis E.J., Messenguy F., Mewes H.W., Miesga T., Mostl D.,
 RA Muller-Auer S., Nentwich U., Obermaler B., Piravandi E., Pohl T.M.,
 RA Portetle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
 RA Schaefer M., Scherens B., Scholler P., Schwager C., Schwarz S.,
 RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhaesele P.,
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hobeisel J.D.;
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII";
 RT Nature 387:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RL EMBL; U17243; AAB67346.1; -.
 DR GDB; S0004292; YLR301W.
 SQ SEQUENCE 244 AA; 27501 MW; 80D813586A1930BB CRC64;

Query Match 2.5%; Score 8; DB 3; Length 244;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 GSTVTSNG 228
 DB 227 GSTVTSNG 234

RESULT 7
 Q9SET1

ID Q9SET1 PRELIMINARY; PRT; 306 AA.
 AC Q9SET1;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Cell wall-plasma membrane linker protein homolog.
 GN CMLP.
 OS *Arabidopsis thaliana* (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20064977; Pubmed=10598107;
 RA Goo J.H., Park A.R., Park W.J., Park O.K.;
 RT "Selection of *Arabidopsis* genes encoding secreted and plasma membrane
 RT proteins";
 RL Plant Mol. Biol. 41:415-423(1999).
 DR EMBL; AF104328; AAD11796.1; -.
 DR HSSP; P24337; IHP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR003882; Pistil_extensin.
 DR InterPro; IPR002965; P_rich_extensn.
 DR Pfam; PF00234; tlyp_alpha_amy1; 1.
 DR PRINTS; PR01217; PRICEXTENSIN.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 306 AA; 31609 MW; F3E6385C50F87827 CRC64;

Query Match 2.5%; Score 8; DB 10; Length 306;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PPTPPV 175
 DB 140 PPTPPV 147

RESULT 8
 ID Q9LIE9 PRELIMINARY; PRT; 334 AA.
 AC Q9LIE9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Similarity to cell wall-plasma membrane linker protein.
 OS *Arabidopsis thaliana* (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA.
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA.
 RX MEDLINE=20363099; Pubmed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of *Arabidopsis thaliana* chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones";
 RL DNA Res. 7:217-221(2000).
 DR EMBL; AF001306; BAB03061.1; -.
 DR HSSP; P24337; IHP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR003882; Pistil_extensin.
 DR InterPro; IPR002965; P_rich_extensn.
 DR InterPro; IPR001768; Tly/amy1_inhbr.

DR Pfam; PF00234; tryp_alpha_aml; 1.
 DR PRINTS; PR01217; PRICEXTENSIN.
 DR PRINTS; PR01218; PSTLEXTENSIN.
 DR SMART; SM00499; AAI; 1.
 SO SEQUENCE 334 AA; 34669 MW; F99159A17AD70524 CRC64;

Query Match 2.5%; Score 8; DB 10; Length 334;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 PTPPPVV 175
 Db 168 PTPPPVV 175

RESULT 9

ID Q9353 PRELIMINARY; PRT; 376 AA.
 AC Q9353;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Cell wall-plasma membrane linker protein.
 GN
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 NCBI_TaxID=3708;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COBRA;
 RA Goodwin W.G., Pallas J.A., Jenkins G.I.;
 RT "Transcripts of a gene encoding a putative cell wall-plasma membrane
 linker protein are specifically cold-induced in Brassica napus."
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X94976; CAA64425.1; -.
 DR HSSP; P24337; IHYP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1_inhbr.
 DR Pfam; PF00234; tryp_alpha_aml; 1.
 DR SMART; SM00499; AAI; 1.
 SO SEQUENCE 376 AA; 38667 MW; 7E12AD838938A4F8 CRC64;

Query Match 2.5%; Score 8; DB 10; Length 376;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 PTPPPVV 175
 Db 175 PTPPPVV 182

RESULT 10

ID Q9UG05 PRELIMINARY; PRT; 88 AA.
 AC Q9UG05;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE BK989H1.1 (Novel protein) (Fragment).
 GN BK989H1.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC Ramsey H.;
 RA Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z83851; CAB62962.1; -.
 FT NON_TER 1 1
 SO SEQUENCE 88 AA; 9699 MW; C2ED8FE7AB67CFA1 CRC64;

Query Match 2.2%; Score 7; DB 4; Length 88;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 PVTAPF 190
 Db 57 PVTAPF 63

RESULT 11

ID Q9CDX3 PRELIMINARY; PRT; 101 AA.
 AC Q9CDX3;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 50S ribosomal protein L24.
 GN RPLX OR IL2088.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Lactococcus.
 NCBI_TaxID=1360;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Mauer S., Jallion O., Malarre K.,
 RA Weisenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 lactis ssp. lactis IL1403."
 RL Genome Res. 11:731-753 (2001).
 DR EMBL; AE006437; AAK06186.1; -.
 DR InterPro; IPR000302; KOW_motif.
 DR InterPro; IPR003256; Ribosomal_L24.
 DR Pfam; PF00467; KOW; 1.
 DR Prodom; PD001677; Ribosomal_L24; 1.
 DR TIGRPFAM; TIGR01079; rplX bact; 1.
 DR PROSITE; PS01108; RIBOSOMAL_L24; UNKNOWN_1.
 DR Ribosome; PS01108; Complete proteome.
 KW RIBOSOMAL_L24; 101 AA; 10877 MW; B4CAF87BE72C9021 CRC64;
 SO SEQUENCE 101 AA; 10877 MW; B4CAF87BE72C9021 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 101;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 VKTGDV 285
 Db 3 VKTGDV 9

RESULT 12

ID Q8S142 PRELIMINARY; PRT; 111 AA.
 AC Q8S142;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE P0042A10.23 protein.
 GN P0042A10.23.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriaristidae; Oryzaceae; Oryza.
 NCBI_TaxID=39947;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RA "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
 clone: P0042A10."
 RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003343; BAB90081.1; -.

SO SEQUENCE 111 AA; 11245 MW; 212977965AB0D90A CRC64;
 Query Match 2.2%; Score 7; DB 10; Length 111;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 PAPVAP 184
 DB 44 PAPVAP 50

RESULT 13
 Q8RUE2 PRELIMINARY; PRT; 122 AA.
 AC Q8RUE2; (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE P0703B11.17 protein (P0485B12.9 protein).
 GN P0703B11.17 OR P0485B12.9.
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0703B11.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa (Japonica cultivar-group) genomic DNA, chromosome 1, PAC
 clone: P0485B12.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF003302; BAB85298.1;
 DR EMBL; AF003348; BAB86479.1;
 DR EMBL; AF003348; BAB86479.1;
 SQ SEQUENCE 122 AA; 13319 MW; D988C108BA8FA070 CRC64;

Query Match 2.2%; Score 7; DB 10; Length 122;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 AGSTVTS 226
 DB 23 AGSTVTS 29

RESULT 14
 Q9ZWT6 PRELIMINARY; PRT; 133 AA.
 AC Q9ZWT6; (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DE 01-MAY-1999 (TREMBlrel. 21, Last annotation update)
 GN ZMGRI1 protein.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LEAF SHEATH;
 RA MEDLINE=99408265; PubMed=10480388;
 RA Ogawa M., Kusano T., Koizumi N., Katsumi M., Sano H.;
 RT "Gibberellin-responsive genes: high level of transcript accumulation
 in leaf sheath meristematic tissue from Zea mays L.";

RL Plant Mol. Biol. 40:645-657 (1999).
 DR EMBL; AB018587; BAA74803.1;
 DR HSSP; P24337; 1HYP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1_inhbr.
 DR Pfam; PF00234; TRY_alpha_amy1; 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 133 AA; 13466 MW; 3B7B3B7CDDC31093 CRC64;

Query Match 2.2%; Score 7; DB 10; Length 133;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PTPPPV 174
 DB 26 PTPPPV 32

RESULT 15
 Q9ST25 PRELIMINARY; PRT; 133 AA.
 AC Q9ST25; (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 GN ZMGRI1 protein.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LEAF SHEATH;
 RA MEDLINE=99408265; PubMed=10480388;
 RA Ogawa M., Kusano T., Koizumi N., Katsumi M., Sano H.;
 RT "Gibberellin-responsive genes: high level of transcript accumulation
 in leaf sheath meristematic tissue from Zea mays L.";
 RL Plant Mol. Biol. 40:645-657 (1999).
 DR EMBL; AB018588; BAA74804.1;
 DR HSSP; P24337; 1HYP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1_inhbr.
 DR Pfam; PF00234; TRY_alpha_amy1; 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 133 AA; 13465 MW; F2AF4E1584FF9631 CRC64;

Query Match 2.2%; Score 7; DB 10; Length 133;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PTPPPV 174
 DB 26 PTPPPV 32

RESULT 16
 Q9FWP4 PRELIMINARY; PRT; 136 AA.
 AC Q9FWP4; (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 GN Putative lipid transfer protein.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONARE;
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
 RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalek H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.,
 RT "Oryza sativa chromosome 10 BAC OSUNBA0015015 genomic sequence."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC026758; AAG13475.1; -
 DR HSSP; P24337; IHP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1_inhbr.
 DR Pfam; PF00234; tryp_alpha_amy1; 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 136 AA; 13982 MW; 9F785B01C3115NAF CRC64;

Query Match 2.2%; Score 7; DB 10; Length 136;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 PTPPVV 164
 DB 37 PTPPVV 43

RESULT 17
 P91181 PRELIMINARY; PRT; 141 AA.
 AC P91181;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 15.0 kDa protein.
 GN C50F2.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Felodetrinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 283:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Du Z., Le T.T.;
 RT "The sequence of C. elegans cosmid C50F2.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission."
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U80445; AAB37800.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 141 AA; 15050 MW; 1F655AC7BAE14E99 CRC64;

Query Match 2.2%; Score 7; DB 5; Length 141;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 STVTSNG 228
 DB 58 STVTSNG 64

RESULT 18
 Q9FWP0 PRELIMINARY; PRT; 142 AA.

AC Q9FWP0;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative lipid transfer protein.
 GN OSUNBA0015015.15.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Euphorbiaceae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONARE;
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
 RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalek H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.,
 RT "Oryza sativa chromosome 10 BAC OSUNBA0015015 genomic sequence."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC026758; AAG13488.1; -
 DR HSSP; P24337; IHP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1_inhbr.
 DR Pfam; PF00234; tryp_alpha_amy1; 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 142 AA; 14393 MW; C3E8AA4F34A63AD1 CRC64;

Query Match 2.2%; Score 7; DB 10; Length 142;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 PTPPVV 164
 DB 37 PTPPVV 43

RESULT 19
 Q8THE9 PRELIMINARY; PRT; 165 AA.
 AC Q8THE9;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Multiple resistance/pH regulation related protein E (Na⁺/H⁺
 DE antiporter).
 GN MKP8 OR M44568.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smitnov S., Ancoor D., Brown A.,
 RA Allen N., Naylor J., Strange-Thomann W., DeAtellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirelli A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Unayam L.A., White O., White R.H., de Maccario E.C.,
 RA Ferry J.G., Jarrell K.F., Ding H., Maccario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE011178; BAM07907.1; -
 KW Complete proteome.
 SQ SEQUENCE 165 AA; 18327 MW; DCAB72532F74E05E CRC64;

Query Match 2.2%; Score 7; DB 17; Length 165;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LGLIFGV 24
DB 29 LGLIFGV 35

RESULT 20

0926C3 PRELIMINARY; PRT; 176 AA.

AC 0926C3; PRELIMINARY; PRT; 176 AA.
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DE Probable peptidoglycan-associated lipoprotein precursor.
GN PAL OR R02738 OR SMC02942.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RA MEDLINE=11396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batur J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Maury D.,
RA Pohl T., Portet-Berle P., Purnelle A., Purnelle B., Rampeger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gilbert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: A591791; CAZ47317.1;
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00661; OmpA, 1.
DR ProDom: PD000930; Bac_OmpA, 1.
KM SEQUENCE 176 AA; 18735 MW; A6F162CB35042268 CRC64;
SQ COMPLETE proteome.

Query Match 2.2%; Score 7; DB 16; Length 176;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 LAGCASK 36
DB 30 LAGCASK 36

RESULT 21

09XW11 PRELIMINARY; PRT; 177 AA.

AC 09XW11; PRELIMINARY; PRT; 177 AA.
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE V54E2A.4 protein.
GN V54E2A.4
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Rhabditinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lloyd C.R.;
RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=99069613; PubMed=9851916;
RX none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";

RL Science 282:2012-2018(1998).
DR EMBL: AL032646; CNA21686.1;
SQ SEQUENCE 177 AA; 20062 MW; E3BA47B50CCFA329 CRC64;

Query Match 2.2%; Score 7; DB 5; Length 177;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 62 STSGSGS 68

RESULT 22

099YU1 PRELIMINARY; PRT; 179 AA.

AC 099YU1; PRELIMINARY; PRT; 179 AA.
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DE Hypothetical protein SPY1538.
GN SPY1538.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RA MEDLINE=1132684; PubMed=1126296;
RA Ferretti J.J., McShane W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Szate S., Suvorov A.N., Kerton S., Lai H.S., Lin S.P.,
RA Olan Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.,
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: A8006586; AAK34332.1;
DR InterPro: IPR004398; Com_hypoth95.
DR InterPro: IPR002052; N6_Mcase.
DR Pfam: PF03602; Com_hypoth95, 1.
DR TIGRFAMs: TIGR00095; Com_hypoth95, 1.
DR PROSITE: PS00092; N6_MTASE; UNKNOWN, 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 179 AA; 19567 MW; BD3A72BFF541137 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 179;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GSGGLAI 58
DB 51 GSGGLAI 57

RESULT 23

09SKR4 PRELIMINARY; PRT; 180 AA.

AC 09SKR4; PRELIMINARY; PRT; 180 AA.
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE ART5. protein (Fragment).
GN ART5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuehl M., Glowacki G., Haag F., Koch-Noite F.,
RT "Molecular cloning and characterization of human ART5, a secretory
RT ecto-mono (ADP-ribosyl) transferase from testis.";
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ294468; CAC07429.1; -
 DR InterPro; IPR000768; ART_family.
 DR Pfam; PF01129; ART; 1.
 FT NON_TER 1 1
 FT NON_TER 180 180
 SO SEQUENCE 180 AA; 20101 MW; 8DC89A96B4B9452 CRC64;

Query Match 2.2%; Score 7; DB 6; Length 180;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 212 RRFGRAT 218
 |||||
 Db 151 RRFGRAT 157

RESULT 24

082FF8 PRELIMINARY; PRT; 189 AA.
 AC 082FF8;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Putative membrane protein.
 GN YPO1754.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia
 NCBI_TaxID=63;

SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebatina M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holtroyd S., Jagsels K., Karlyshev A.V.,
 RA Leather S., Moul S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 DR EMBL; AJ414150; CAC90573.1; -
 DR InterPro; IPR003810; DUF204.
 DR Pfam; PF02659; DUF204; 2.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 189 AA; 20660 MW; 598D4DF73D5E48C5 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 189;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 GLIRGVI 25
 |||||
 Db 40 GLIRGVI 46

RESULT 25

09V3T8 PRELIMINARY; PRT; 195 AA.
 AC 09V3T8; Q9VK86;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE SC35 protein (Lb22469p).
 GN SC35 OR CG5442.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;

Qy 11
 |||||
 Db 11
 SEQUENCE FROM N.A. (LONG ISOFORM).
 RP

RA Bourbon H.M., Allemand E., Soret J., Tazi J.;
 RT "Characterization of the Drosophila ortholog of the mammalian pre-mRNA
 RT splicing factor SC35."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

SO SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).

RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mills G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Panicoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlehnina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svateks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).

SO SEQUENCE FROM N.A.

RP STRAIN=BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.

DR EMBL; AF232775; AAF43415.1; -
 DR EMBL; AB003636; AAF51192.1; -
 DR EMBL; AB003636; AAF51193.1; -
 DR EMBL; AY069584; AAL39729.1; -
 DR HSSP; P11940; 1CVJ.
 DR Flybase; FBgn040286; SC35.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 KW Alternative splicing
 FT VARSPLIC 1 83 MISSING (IN SHORT ISOFORM).
 SO SEQUENCE 195 AA; 21400 MW; 65E922F8C34C7536 CRC64;

Query Match 2.2%; Score 7; DB 5; Length 195;

Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 51 SSGCGLA 57
Db 179 SSGCGLA 185

RESULT 26

OBXXR6 PRELIMINARY; PRT; 205 AA.

AC 08XXR6; 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Hypothetical protein RSC2047.
GN RSC2047 OR RS03612.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_Taxid=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM1100;
RA MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlet M., Billault A., Brotilier P., Camus J.C., Catroliello L.,
RA Chandler M., Choiane N., Claudel-Renard C., Cumac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Weisenbach J., Bouchet C.A., Whalen M., Wincker P., Levy M.,
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RT Nature 415:497-502 (2002).
DR EMU; A1646068; CAD15754.1; -
DR InterPro; IPR002667; IPR_1isomerase.
DR InterPro; IPR000086; NUDIX_Hydrolase.
DR Pfam; PF00293; NUDIX; 1.
DR PRINTS; PR00502; NUDIXFAMILY.
DR Prodom; PD004109; IPR_1isomerase; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 205 AA; 22905 MW; 011998A4555B1B10 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 205;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 27

OB1130 PRELIMINARY; PRT; 208 AA.

AC 081130; 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CCAT-box binding factor HAP3 Homolog.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WS-0;
RA MEDLINE=96319234; PubMed=9657152;
RA Loezan T., Ohto K.M., West M.A., Lo R., Kwong R.W.,
RA Yanagishi K., Fischer R.L., Goldberg R.B., Harada J.J.;
RT "Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo
development in vegetative cells.";
RL Cell 93:1195-1205 (1998).

DR EMBL; AF036684; AAC39488.1; -
DR HSSP; P19267; 1A7W.
DR InterPro; IPR003958; CBFA_NFYB_domain.
DR InterPro; IPR003957; CBFA_NFYB_topis.
DR InterPro; IPR004822; Histone_core.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
SQ SEQUENCE 208 AA; 22693 MW; 4E2D249AE325DDB CRC64;

Query Match 2.2%; Score 7; DB 10; Length 208;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 314 YVDPPLTV 320
Db 105 YVDPPLTV 111

RESULT 28

O9SPD8 PRELIMINARY; PRT; 208 AA.

AC O9SPD8; 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE T26P17.20
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bel O., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federpriel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T26P17 from chromosome
I.";
RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RN Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bel O., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Davis R., Federpriel N., Theologis A.,
RA Ecker J.;
RN Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bel O., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federpriel N.,
RA Theologis A., Ecker J.;
RN Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
DR EMBL; AC013482; AAF16537.1; -
DR HSSP; P19267; 1A7W.
DR InterPro; IPR003958; CBFA_NFYB_domain.
DR InterPro; IPR003957; CBFA_NFYB_topis.
DR InterPro; IPR004822; Histone_core.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.

DR PRINTS, PRO0615; CCAATSUBUNTA.
 SQ SEQUENCE 208 AA; 22679 MW; 0DEB469D9FB8EFDCC CRC64;
 Query Match 2.2%; Score 7; DB 10; Length 208;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
 |||||
 Db 105 YVDPLTV 111

RESULT 29
 Q9S2U3 PRELIMINARY; PRT; 211 AA.
 AC Q9S2U3;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE Putative membrane protein.
 GN SC02042 OR SC466.11C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 NCBI_TaxID=1902;
 RX STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Frazer A., Goble A., Hidalgo J., Hornaby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovitch B., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL; AL096884; CAB51434.1; -
 SQ SEQUENCE 211 AA; 21223 MW; 4A8123837C5683B8 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 211;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GTATVAG 221
 |||||
 Db 50 GTATVAG 56

RESULT 30
 Q9FYQ4 PRELIMINARY; PRT; 212 AA.
 AC Q9FYQ4;
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE Hypothetical protein (P0011G08.31 protein).
 GN P0011G08.31.
 OS Oryza sativa (rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 RX STRAIN=CV, NIPPONBAR;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RA "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0433F09.";
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, NIPPONBAR;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0011G08.31";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002539; BAB08181.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 212 AA; 23424 MW; C5B5EB20D459DA24 CRC64;

Query Match 2.2%; Score 7; DB 10; Length 212;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 SSGSGSHR 49
 |||||
 Db 160 SSGSGSHR 166

RESULT 31
 Q9VT29 PRELIMINARY; PRT; 213 AA.
 AC Q9VT29;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE CG14414 protein.
 GN CG14414.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RX STRAIN=BERKELEY;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Klammer B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lakso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlehn N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003496; AAF48376.2; -
 DR FLYBASE; FBgn0030571; CG14414.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS00102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 SQ SEQUENCE 213 AA; 23408 MW; 72D078316A76DF7A CRC64;
 Query Match 2.2%; Score 7; DB 5; Length 213;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 50 TSGSGGL 56
 DB 7 TSGSGGL 13
 RESULT 32
 OGSN98 PRELIMINARY; PRT; 216 AA.
 AC OGSN98;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE RNA polymerase subunit.
 GN RPOA.
 OS Euglena gracilis.
 OG Chloroplast.
 OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
 OX NCBI_TaxID=3039;
 RX STRAIN=Z;
 RP SEQUENCE FROM N.A.
 RC MEDLINE=93347989; PubMed=8346031;
 RA Hallick R.B., Hong L., Dreger R.G., Favreau M., Monfort A., Orsat B.,
 RA Spielmann A., Stutz E.,
 RT "Complete sequence of *Euglena gracilis* chloroplast DNA,"
 RL Nucleic Acids Res. 21:3537-3544(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94198895;
 RA Stevenson J.K., Hallick R.B.,
 RT "The psal operon pre-mRNA of the *Euglena gracilis* chloroplast is
 RT processed into photosystem I and II mRNAs that accumulate
 RT differentially depending on the conditions of cell growth,"
 RL Plant J. 5:247-260(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RX MEDLINE=95047317; PubMed=7958642;
 RA Hong L., Hallick R.B.,
 RT "A group III intron is formed from domains of two individual group II
 RT introns,"
 RL Genes Dev. 8:1589-1599(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RA Schluegger B., Stutz E.,
 RT "The *Euglena gracilis* chloroplast genome: structural features of a DNA
 RT region possibly carrying the single origin of DNA replication,"
 RL Curr. Genet. 8:629-634(1984).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RX MEDLINE=94250697; PubMed=8193167;
 RA Orsat B., Spielmann A., Marc-Martin S., Lemberger T., Stutz E.,
 RT "Analysis of the 22 kbp long psbd-psbc gene cluster of *Euglena*

RT *gracilis* chloroplast DNA: evidence for overlapping transcription units
 RT undergoing differential processing,"
 RL Biochim. Biophys. Acta 1218:75-81(1994).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RX MEDLINE=95007823; PubMed=7923415;
 RA Hong L., Hallick R.B.,
 RT "Gene structure and expression of a novel *Euglena gracilis* chloroplast
 RT operon encoding cytochrome b6 and the beta and epsilon subunits of the
 RT H(+)-ATP synthase complex,"
 RL Curr. Genet. 25:270-281(1994).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RX MEDLINE=96079990;
 RA Jenkins K.P., Hong L., Hallick R.B.,
 RT "Alternative splicing of the *Euglena gracilis* chloroplast roaA
 RT transcript,"
 RL RNA 1:624-633(1995).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RA Hallick R.B.,
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X70810; CAC69148.1; -
 KW Chloroplast.
 SQ SEQUENCE 216 AA; 25364 MW; BFBF1385921BC187 CRC64;
 Query Match 2.2%; Score 7; DB 8; Length 216;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 12 OKPIKRL 18
 DB 165 OKPIKRL 171
 RESULT 33
 O29808 PRELIMINARY; PRT; 225 AA.
 AC O29808;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein AF0441.
 GN AF0441.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrtides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Adams M.D., Loftus B.,
 RA Kitzes E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Ketch C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,
 RA Overbeek R., Goehring T.D., Weidman J.F., McDonald L., Uterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujita C., Garland S.A.,
 RA Macon T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.,
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*,"
 RL Nature 390:364-370(1997).
 DR EMBL; AE001074; AAB90793.1; -
 DR TIGR; AF0441; -
 DR InterPro; IPR001989; Radical_activat.
 DR Pfam; PF02143; Radical_activat; 1.

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 225 AA; 25873 MW; 27A28FB3CCE2188A CRC64;

Query Match 2.2%; Score 7; DB 17; Length 225;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 GDLKVR 124
Db 123 GDLKVR 129

RESULT 34

Q92KU4 PRELIMINARY; PRT; 226 AA.

AC 092KU4; 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein R00142.
GN R00142 OR SMC04110.

OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
[1]

SEQUENCE FROM N.A.

RC STRAIN=1021.
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,
RA Godre T., Goffeau A., Kahn D., Kiss E., Lelaure V., Maury D.,
RA Pohl T., Portetelle P., Puehler A., Purnelle B., Ransperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591782; CAC41529.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 226 AA; 24278 MW; CF261B591D7B48E CRC64;

Query Match 2.2%; Score 7; DB 16; Length 226;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 LAGCASK 36
Db 20 LAGCASK 26

RESULT 35

Q9H5P8 PRELIMINARY; PRT; 228 AA.

AC 09H5P8; 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Homo sapiens (human).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OC NCB1_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.

RP TISSUE=LUNG;

RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA Iwagishi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Tsugei T., Sugano S.,
RT "NEO human cDNA sequencing project."
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026842; BAB15572.1; -.

SQ SEQUENCE 228 AA; 26292 MW; 171D0DCD1364E01F CRC64;

Query Match 2.2%; Score 7; DB 4; Length 228;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 QSSRPV 149
Db 166 QSSRPV 172

RESULT 36

Q9KLV7 PRELIMINARY; PRT; 241 AA.

AC 09KLV7; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein VCA0634.
GN VCA0634.

OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
[1]

SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey B.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Esmolaeva M.D., Vamathevan J., Bess S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uetzerack T., Fleischmann R.D., Nieman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004393; AAF96535.1; -.
DR TIGR; VCA0634; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 241 AA; 27259 MW; 1AE3F947FCD04B27 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 241;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 SKIAQRY 88
Db 192 SKIAQRY 198

RESULT 37

Q8TUT2 PRELIMINARY; PRT; 241 AA.

AC 08TUT2; 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cobalamn-5-phosphate synthase.
GN COBS OR MK1671.

OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
[1]

RP SEQUENCE FROM N.A.

RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polunin N.N.,
RA Shcherbina O.V., Shakova V.V., Belova G.I., Mavrin N.N.,
RA Natale D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Stetter K.O.,
RA Malyn A.G., Koonin E.V., Kozlyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19

```

RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AB010456; AM02884.1; -.
KW Complete proteome.
SQ SEQUENCE 241 AA; 25264 MW; E1C25302FB7FEB6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 17; Length 241;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 54 GGLAIGS 60
Db 109 GGLAIGS 115

RESULT 38
O9HNY3 PRELIMINARY; PRT; 249 AA.
ID O9HNY3;
AC O9HNY3;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE ABC transport protein.
GN TRP2 OR UNG1893G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxId=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2050448; PubMed=11016950;
RA NG W.V., Kennedy S.P., Mahaitas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Laakey S.R., Baliga N.S., Thoreson V., Sirogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weiti R., Goo Y.A.,
RA Leitauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Iammarig T.A., Peck R.F., Pohlschoder M., Spudich J.L., Jung K.-H.,
RA Alami M., Fretas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Benhart H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005088; AAC20087.1; -.
DR InterPro; IPR003439; AAA_Atpase.
DR InterPro; IPR003439; ABC_transport.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transport; 1.
DR SMART; SM00382; AAA; 1.
KW Complete proteome.
SQ SEQUENCE 249 AA; 26679 MW; 06F96H1ABD2AE19 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 17; Length 249;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 215 GTATVAG 221
Db 60 GTATVAG 66

RESULT 39
O51373 PRELIMINARY; PRT; 259 AA.
ID O51373;
AC O51373;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Hypothetical protein BB0412.
GN BB0412.
OS Borrelia burgdorferi (Lyme disease spirochete).
OS Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId=139;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Castens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kervatage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utecht S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
DR EMBL; AB01146; AAC6791.1; -.
DR TIGR; BB0412; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 30809 MW; 2137010F9140DOFF CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 259;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 99 NNINSSY 105
Db 109 NNINSSY 115

RESULT 40
O94238 PRELIMINARY; PRT; 269 AA.
ID O94238;
AC O94238;
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Hypothetical 30.2 kDa protein.
GN F55A4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Rhabditinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Bradshaw H.;
RT "The sequence of C. elegans cosmid F55A4."
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U67949; AB07562.1; -.
DR HSSP; Q91836; 1D12.
DR InterPro; IPR001159; DS_RBD.
DR Pfam; PF00035; dsm; 2.
DR SMART; SM00358; DSRM; 2.
DR PROSITE; PS0137; DS_RBD; 2.
KW Hypothetical protein.
SQ SEQUENCE 269 AA; 30169 MW; 2028PE3245563191 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 5; Length 269;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy	315	VDP	TVL	321
Db	14	VDP	TVL	20

Search completed: July 8, 2003, 11:20:55
Job time : 111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 3, 2003, 13:47:57 ; Search time 2083 seconds
(without alignments)
4498.849 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670
Sequence: 1 MTVTAINSONOKPIKRLGL.....LFEFRISNGVYVDLTVLK 322

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODL=frame_plus_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_SPOOL/US10018706/runat_30062003_091104_23794/app_query.fasta_1.519
-DB=genemb1 -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45
-DOCMATCH=200 -THR_SCORE=DCT -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=2000000000
-USER=US10018706@cgn2.1.1.1687@runat_30062003_091104_23794 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_secs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_secs:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1670	100.0	969	6 AX063563	AX063563 Sequence
2	1670	100.0	48328	6 AX067452	AX067452 Sequence
3	1666	99.8	966	6 AX063565	AX063565 Sequence
4	298	17.8	2910	1 PSEBPOS	D26134 Pseudomonas
5	298	17.8	13031	1 AB004782	AB004782 Pseudomonas
6	289	17.3	58996	1 AB034704	AB034704 Ruditiviva
7	286	17.1	3122	1 AB006073	AB006073 Pseudomonas
8	282	16.9	5876	1 AF421351	AF421351 Acetobact
9	253	15.1	10026	1 AB003925	AB003925 Xylella f
10	249.5	14.9	12956	1 AB006198	AB006198 Pasteurel
11	245	14.7	3539	1 PPY19122	Y19122 Pseudomonas
12	238	14.3	9720	1 AE000369	AE000369 Escherich
13	238	14.3	55175	1 EC028375	U28375 Escherichia
14	238	14.3	26668	1 AP002563	AP002563 Escherich
15	234.5	14.0	10773	1 AB012272	AB012272 Xanthomon
16	234	14.0	10622	1 AB004905	AB004905 Pseudomon
17	232.5	13.9	22400	1 AE008839	AE008839 Salmonell
18	232.5	13.9	230050	1 AL627277	AL627277 Salmonell
19	231.5	13.9	1552	1 AF260132	AF260132 Pseudomon
20	227.5	13.6	2609	1 AF244357	AF244357 Coxilla
21	225	13.5	9360	1 AE005516	AE005516 Escherich
22	222.5	13.3	207050	1 AL646063	AL646063 Ralstonia
23	219.5	13.1	11435	1 U32753	U32753 Haemophilus
24	219.5	13.1	11555	1 AE011804	AE011804 Xanthomon
25	219.5	13.1	14193	1 AE002061	AE002061 Deinococc
26	218.5	13.1	10592	1 AE013686	AE013686 Yersinia
27	218.5	13.1	220050	1 AJ414156	AJ414156 Yersinia
28	213.5	12.8	1433	1 EC06CMTLPD	L07669 Escherichia
29	213.5	12.8	10573	1 AE005502	AE005502 Escherich
30	213.5	12.8	11457	1 AE000358	AE000358 Escherich
31	213.5	12.8	72221	1 EC029579	U29579 Escherichia
32	213.5	12.6	270365	1 AP002562	AP002562 Escherich
33	210	12.6	274050	1 AL627276	AL627276 Salmonell
34	209	12.5	1290	1 SDU6131	AJ006131 Salmonell
35	209	12.5	2179	6 AR210638	AR210638 Sequence
36	209	12.5	20984	1 AE008833	AE008833 Salmonell
37	208.5	12.5	10104	1 AE004139	AE004139 Vibrio ch
38	206	12.3	1612	1 ECORPOS	D17549 Escherichia
39	205	12.3	12362	1 AB009547	AB009547 Neisseria
40	203	12.2	329861	1 NMA5522491	ALI62756 Neisseria
41	203	12.2	349619	1 AP002996	AP002996 Mesorhizo
42	198	11.9	11085	1 AE0002497	AE0002497 Neisseria
43	198	11.9	349980	6 AX044032	AX044032 Sequence
44	198	11.9	349980	6 AX044033	AX044033 Sequence
45	197.5	11.8	3692	12 EC0FV1AB	Z21706 E. coli/C.fr

RESULT 1

ALIGNMENTS

AX063563
 LOCUS AX063563 969 bp DNA linear PAT 24-JAN-2001
 DEFINITION Sequence 1 from Patent WO0100838.
 ACCESSION AX063563
 VERSION AX063563.1 GI:12541291
 KEYWORDS
 SOURCE Moraxella catarrhalis.
 ORGANISM Moraxella catarrhalis.
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 Moraxella.

REFERENCE
 1 (bases 1 to 969)
 AUTHORS Thomare, J.S.
 TITLE Cloning of Baelb10 antigen from moraxella (branhamella) catarrhalis
 JOURNAL Patent: WO 0100838-A 1 04-JAN-2001;
 SMITHKLINE BEECHAM BIOLOGICALS (S.A.)
 FEATURES
 source 1..969
 Location/Qualifiers
 /organism="Moraxella catarrhalis"
 /db_xref="taxon:480"

BASE COUNT 265 a 214 c 242 g 248 t

ORIGIN
 Alignment Scores:
 Pred. No.: 1.07e-90 Length: 969
 Score: 1670.00 Matches: 322
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-10-018-706-2 (1-322) x AX063563 (1-969)

QY 1 MetThValThrIleAlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeu 20
 DB 1 ATGACCTGACGATGACATTCATTCACAAATCCAAACCCATCAGCATTCGGCTTG 60

QY 21 IlePhgGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40
 DB 61 ATTTTGGTGTGATCACACCTGATTTGGCAGAGATGCCAGTAACTTAAT 120

QY 41 SerThSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySer 60
 DB 121 AGTACCTCAGGTCGGGCGACTCATCTCAGTTCAGGTCAGTGGTTGGCAATAGGTTCA 180

QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValLysGlnGlyAspThr 80
 DB 181 CAGGTATATCCGACAGTACGAGCGGTACAAATCGCTATCAGGTGAAGCAGGCGATACT 240

QY 81 ValSerLysIleAlaGlnArgTyrGlyLeuAsnTyrArgLuiIleGlyHisIleAsnAsn 100
 DB 241 GTCAGTAAATTCCTCAGCTTATGATTAATTTGGCGTGAATGGACACATTAAATAT 300

QY 101 LeuAsnSerSerTyrThrIleTyrThrGlyGlnTyrLeuThrLeuTyrSerGlyAspLeu 120
 DB 301 CTAAATATGACGATTAACGATTATACAGTCATACGTCGATTAATGTCAGGATCTC 360

QY 121 LysValArgGlyArgSerIleSerSerGlyValAsnThrLysThrProSerProVal 140
 DB 361 AAGTCGTGAGGCTATCATCAGCTCTGGTGTGATTCAGTCTCACACCTTCGCTGTG 420

QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThrPro 160
 DB 421 GCGGTTCAGTCAACGACACCACTACAGACAGATCTCCGCTACAAAACCCACGCCA 480

QY 161 ProValValValValLysLysProThrProThrProProValValGlnGlnProAlaPro 180
 DB 481 CCGTGTGTGTGTAAACCAACCCACACGCTCCGCTGTGTGTGCGACGCCACCA 540

QY 181 ValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValIleGlnPhe 200
 DB 541 GTTGCCCAACCAATGACAGAAACCACTTTGGCCAGGTGATCAGGCGGTGATGCAATTT 600

QY 201 ArgTyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAla 220

DB 601 CGGTATCTCTTGTTGGTGGACCAATCCAGTGGTTCAGCGCTTGGTACCGGACAGTGGCC 660

QY 221 GlySerThrValThrSerAsnGlyMetTyrPheSerGlyArgAspGlyAspLysIleAsn 240
 DB 661 GCGTCACTGTTACCGATTAATGCAATGCTGTTTTCGAGAGATGGCCATTAAATTAAC 720

QY 241 AlaSerAspAlaGlyThrValIleGlnAlaAspHisAsnMetArgGlyValAsnIleVal 260
 DB 721 GCCAGTATCAGGACACAGCATTCATTCAGCTGATCAACATTCAGACGGCGGATTTGTG 780

QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
 DB 781 ATTCAGCATACCAATGATGATTTGTTCAAGCTATATCCATTAAGAGCGCTCAAGTTAAA 840

QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGly 300
 DB 841 ACAGGCGATACGGTGGTACCGGTACCGCTATTGCAAGCAATGAAGAAATCCAGCAAGCGGT 900

QY 301 AlaAlaLeuPheGlyLupheArgIleSerArgAsnGlyValTyrValAspProLeuThrVal 320
 DB 901 GCGGCACTATTGAAATTTAGAAATTTCTAGAAATGGCGTGTATGATCCATTGACAGTA 960

QY 321 LeuLys 322
 DB 961 CTTAAA 966

RESULT 2
 AX067452 48326 bp DNA linear PAT 24-JAN-2001
 LOCUS AX067452
 DEFINITION Sequence 27 from Patent WO0078968.
 ACCESSION AX067452
 VERSION AX067452.1 GI:12545072
 KEYWORDS
 SOURCE Moraxella catarrhalis.
 ORGANISM Moraxella catarrhalis.
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 Moraxella.

REFERENCE
 1 (bases 1 to 48328)
 AUTHORS Lagace, R.E., Patterson, C. and Berg, K.L.
 TITLE Nucleotide sequences of moraxella catarrhalis genome
 JOURNAL Patent: WO 0078968-A 27 28-DEC-2000;
 Incyte Genomics, Inc. (US)
 FEATURES
 source 1..48328
 Location/Qualifiers
 /organism="Moraxella catarrhalis"
 /db_xref="taxon:480"

BASE COUNT 14211 a 9322 c 10802 g 13992 t

ORIGIN
 Alignment Scores:
 Pred. No.: 8.02e-89 Length: 48328
 Score: 1670.00 Matches: 322
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-10-018-706-2 (1-322) x AX067452 (1-48328)

QY 1 MetThValThrIleAlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeu 20
 DB 40850 ATGACCTGACGATGACATTCATTCACAAATCCAAACCCATCAGCATTCGGCTTG 40909

QY 21 IlePhgGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40
 DB 40910 ATTTTGGTGTGATCACACCTGATTTGGCAGAGATGCCAGTAACTTAAT 40969

QY 41 SerThSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySer 60
 DB 40970 AGTACCTCAGGTCGGGCGACTCATCTCAGTTCAGTGGTGGCAATAGGTTCA 41029

QY 61 GlnValIleThrAspSerIleGlnGlyValProAsnArgTyrGlnValIleGlnGlySerThr 80
DB 41030 CAGGTTATACCGAGACAGTCAAGGCGTACCAATCGCTATCAGGTGAAGCAGGCGATCT 41089
QY 81 ValSerIleIleAlaGlnArgTyrGlyLeuSerThrArgGluIleGlyHisIleAsnAsn 100
DB 41090 GTTCAGTAAGATTGCTCAGCGTTATGATTAATTCAGTGCATGAGTGAACACATTAAT 41149
QY 101 LeuAsnSerSerTyrThrIleTyrThrGlnGlnIleThrLeuThrLeuThrSerGlyAspLeu 120
DB 41150 CTAAATAGAGATTATGATTATACAGGCAATGCTGACTTATGCTGATGCTGATGCTC 41209
QY 121 LysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProVal 140
DB 41210 AAGTCCGAGCGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 41269
QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThrPro 160
DB 41270 GCGGTTCAGTCAAGCAGACCAAGACAGACATCTGCTGCTGCTGCTGCTGCTGCTG 41329
QY 161 ProValValValValLysLysProThrProThrProProValValGlnGlnProAlaPro 180
DB 41330 CCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 41389
QY 181 ValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
DB 41390 GTTGCCCAACCAAGTCAAGACCAATTCGCAAGGCTGCTGCTGCTGCTGCTGCTG 41449
QY 201 ArgTyrProValGlyValAlaThrAsnProValValArgArgPheGlyThrAlaThrValAla 220
DB 41450 CCGTATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 41509
QY 221 GlySerThrValThrSerAsnGlyMetTyrPheSerGlyValArgAspGlyAspLeuIleAsn 240
DB 41510 GCGTCAACGTTACCGATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 41569
QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleVal 260
DB 41570 GCCAGTAAGCAGGACCAATCTTCAAGTCAATCAATATGACACCGGCGGATATTG 41629
QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
DB 41630 ATTACAGCATACCAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 41689
QY 281 ThrGlyAspThrValArgThrGlyGlnAlaGlyIleAlaSerMetLysAsnGlnProSerGly 300
DB 41690 ACAGGCGATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 41749
QY 301 AlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyrValAspProLeuThrVal 320
DB 41750 GCGGCACTATTGTAATTGTAATTTCTAGAAATGGCGTGTATGTATCATTCATGAC 41809
QY 321 LeuLys 322
DB 41810 CTTAA 41815

RESULT 3
AX063565
LOCUS AX063565 966 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 3 from Patent WO0100838.
ACCESSION AX063565
VERSION AX063565.1 GI:12541292
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
1 (bases 1 to 966)
REFERENCE
AUTHORS Thomard, J.S.
TITLE Cloning of babB10 antigen from moraxella (branhamella) catarrhalis
JOURNAL Patent: WO 0100838-A 3 04-JAN-2001;
SMITHKLINE BEECHAM BIOLOGICALS (S.A.)

FEATURES
Source Location/Qualifiers
1..966
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 263 a 214 c 242 g 247 t
ORIGIN
Alignment Scores:
Pred. No.: 1,85e-90 Length: 966
Score: 1666.00 Matches: 321
Percent Similarity: 99.69% Conservative: 0
Best Local Similarity: 99.69% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: Gaps: 0
US-10-018-706-2 (1-322) x AX063565 (1-966)
QY 1 MetThrValThrIleAlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeu 20
DB 1 ATGACTGTACGATACCAATCAATTCACAAATCAAAACCCATCAAGCATGGGCTTG 60
QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40
DB 61 ATTTTGGTGTATCCACTTCGATTTGGCAGATGTGCCAGTAAAGCCAACTTAAT 120
QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySer 60
DB 121 AGTACTCAGGTTCCGAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValIleGlnGlySerThr 80
DB 181 CAGGTTATACCGAGACAGTCAAGGCGTACCAATCGCTATCAGGTGAAGCAGGCGATCT 240
QY 81 ValSerIleIleAlaGlnArgTyrGlyLeuSerThrArgGluIleGlyHisIleAsnAsn 100
DB 241 GTCAAGTAAGTCTCAGCGTTATGATTAATTCGCGAGATGAGATGACATTAAT 300
QY 101 LeuAsnSerSerTyrThrIleTyrThrGlnGlnIleThrLeuThrLeuThrSerGlyAspLeu 120
DB 301 CTAAATAGCGGTTATCGATTATACAGTCAATGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 121 LysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProVal 140
DB 361 AAGTCCGAGCGTGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThrPro 160
DB 421 GCGGTTCAGTCAAGCAGACCAAGACAGACATCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 161 ProValValValValLysLysProThrProThrProProValValGlnGlnProAlaPro 180
DB 481 CCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 540
QY 181 ValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
DB 541 GTTGCCCAACCAAGTCAAGACCAATTCGCAAGGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 201 ArgTyrProValGlyValAlaThrAsnProValValArgArgPheGlyThrAlaThrValAla 220
DB 601 CCGTATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660
QY 221 GlySerThrValThrSerAsnGlyMetTyrPheSerGlyValArgAspGlyAspLeuIleAsn 240
DB 661 GCGTCACTGTTCACGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleVal 260
DB 721 GCCAGTAAGCAGGACCAATTCATTAAGCTGATCAATATGACACCGGCGGATATTG 780
QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
DB 781 ATTACAGCATACCAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 840

Db 891 GGGTGGCATGGCTCGAGCGGTACT-----CTGATCGCGCTTT----- 932
Qy 219 VALAAGlySerThrValThrSerAnglyMetTrpPheSerGlyArgAspGlyAspLeu 238
Db 933 ---GCTCAACGGAAGTTGAATAAGGATGATATAGCCCGCTCAATTGGCCAGCCT 989
Qy 239 ILeAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer 258
Db 990 GTTCCTGGCTGGCTGGTGGGACCTGTGATACCGCGGTAGTGGTTGGCGGGCTACGGC 1049
Qy 259 -----IleValIleGlnHisThrAsnGlyPheValSerSerTrpIleHisIleLysAsp 276
Db 1050 GAGTTGTCATCTCATCAACACACAGACCTACGTGATGCTCAACCGCAG 1109
Qy 277 AlAGlnVallyThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsn 296
Db 1110 CTGCTGTGTGGGGAAGGCAACAGGTCAAGTAGGCGCAATTCGATCCGAGATGGCCTCC 1169
Qy 297 GlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyValAsp 316
Db 1170 AAGAGAACCGATCGGGTGAAGCTGCACTTGAGATTGCGCCGAGGTAAAGCTGTGAT 1229
Qy 317 ProLeuThrValLeu 321
Db 1230 CCACTGCAATATTG 1244
RESULT 5
AE004782/c 13031 bp DNA linear BCT 30-AUG-2000
LOCUS pseudomonas aeruginosa PA01, section 343 of 529 of the complete
DEFINITION genome.
ACCESSION AE004782 AE004091
VERSION AE004782.1 GI:9949772
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 13031)
AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lilm,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
NATURE 406 (6799), 959-964 (2000)
JOURNAL
MEDLINE 2043737
PUBMED 10984043
REFERENCE 2 (bases 1 to 13031)
AUTHORS Stover,C.K., Pham,X.Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Lardig,K., Lilm,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,
Hanock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-May-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
TITLE
JOURNAL Location/Qualifiers
FEATURES
source 1..13031
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
72..1142
/gene="PA3615"
72..1142
/gene="PA3615"
/codon_start=1
/transl_table=11

gene /product="hypothetical protein"
CDS /protein_id="AAG07003.1"
/db_xref="GI:9949773"
/translation="MPEYADYDLMVITTAGGGIMAAHREAGLENSLGFNITLPEFOHANTVDG
VALARLQADLRNRKDAKIMLTQTLREKMSASVLEQFRKRRTVTFGSAKTPVEHP
MPTTTPMAQADLRNRKDAKIMLTQTLREKMSASVLEQFRKRRTVTFGSAKTPVEHP
SCNLSLHFFPFLKFLFVKEADALVLCPPGFGLDEALVTLVQCKSPVLYVILD
OPGGRVHEALFEMOLENNHYILPCPGFGLDEALVTLVQCKSPVLYVILD
KQTPVRLNHLNLEALNHLHFPALSLGSGFQQQAYAEQDEDEFPNLTFLAVF
NRDQRLREILDIYNLPENNMD"
complement (1164..1625)
/gene="PA3616"
complement (1164..1625)
CDS /protein_id="AAG07004.1"
/db_xref="GI:9949774"
/translation="MAIVADTPVAVARAMDLLARERHGAELSRKLROGASAEILD
PALDRLAERGLDESRYLESFTASRARGHPLRIEELAORGLPADIERALGACEV
DMSAQRLBYWRKFPARLPDAREKAOQGFLLAVRGISMSISRLNGRBD"
complement (1631..2671)
/gene="reca"
/note="PA3617"
complement (1631..2671)
CDS /codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG07005.1"
/db_xref="GI:9949775"
/translation="MDENKRLAALAAIGIEROFGGAWRWGDHEROALPAITSGSL
GIDIALGIGLPRKGRIVEIYGPSSGKTLITLSVAREAKQCATCFVPAERALDDY
AKLGVNVDLLVSQDTEQALLETIDMLVRNANADVILVDSVAALVPAELLEGEND
AHVGLQARLMSQALRKITGNINANCLVFINQIRKIKTGMENPETTTGNAIKFYA
SVRLDIRRTGAVEGDEVSETRVAVNKKVSPPRQAEFOILYKGIYRGEIIL
GVOLGIVEKGAWYQSGKIQGKANNKYLEDNBEISVLEKTIIRDQLAKSGPVK
ADAEVADAEAD"
complement (2805..3311)
/gene="PA3618"
complement (2805..3311)
CDS /gene="PA3618"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG07006.1"
/db_xref="GI:9949776"
/translation="MTLMSDLTALAARGLSEILRYLGEQVSTAESCCTGGGIAEATRI
PGSGWFEAGVTVNSRQKTLQLEVESEFPFPAVGAVSREVVEAMVGAORHSGARFAY
AVSGVAGPDGSGPEKRVGTWLMADGERLVSERQCFNDRDRAVRRQVATALTGLLR
MSNGENPQ"
3459..4466
/gene="PA3619"
3459..4466
/gene="PA3619"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07007.1"
/db_xref="GI:9949777"
/translation="NNIAPLKYSLSLALSLSPVASLPAHAACNTALALOPASSLPAL
GGRLVYHSMYEWGDSNNLYLHDFKSKRLOINOGKNEIDEMNAPHSFPGGYLTWPG
RONGAHRYPMAIGCTQAPSNLTAIAGNENPKPSFDRQVLYLKHEDITRLATLVFN
GDGSGVSAKRAVTDGWSSTESMPFLTPSGKRVVYATAGSLRVRRNLRRNGQCA
LATPAAGRDYYPVVRDYAYPLSRTOPAGNDQLAMVAVNSPGRPALPLNHCCQDN
SDAAVVEDYLIFFSTSFDTYSLLDGIAARVVRDPAQINLPDGRQKLGASYTAA
R"
4592..7159
/gene="mutS"
/note="PA3620"
4592..7159
/gene="mutS"


```

Oy 259 -----11eVal111eGlnHisThrAnglypheValSerSerTyrl1eHsile1eysAp 276
Db 9291 GAGTTGATCATCAAAACACAGAGAGCTACGTCAGCTCAACGCAACCCAGG 9232
Oy 277 AAGinValGlyThrG1yAspThrValAagThrG1yGlnArg11eHsMet1ySaen 296
Db 9231 CTGCTGTCGCGGAGGAGGACAGGTCAGGAGGCAATGCGCCAGAGTGGGCTCC 9172
Oy 297 GlnProSerG1yAlaAlaLeuPheGluPheArg11eSerArgAsn1yValTyrlAaP 316
Db 9171 ACAGAGACCGATCGGAGTGAAGCTGCACTTCGAGATTCGCCGCCAGGTAAGCTTCGAT 9112
Oy 317 ProLeuThrVal1eLeu 321
Db 9111 CCACCTGCAATATTTC 9097

RESULT 6
AB034704 58996 bp DNA linear BCT 12-APR-2000
LOCUS Rubriivivax gelatinosus photosynthetic gene cluster, complete cds,
DEFINITION AB034704.1 GI:7416771
ACCESSION AB034704.1 GI:7416771
VERSION
KEYWORDS
SOURCE
ORGANISM
Rubriivivax gelatinosus (strain:IL144) DNA, clone:pgc#12, pgc#6.
Rubriivivax gelatinosus
Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
Rubriivivax.
1 (sites)
Nagashima, K.V., Shimada, K. and Matsuura, K.
Phylogenetic analysis of photosynthetic genes of Rhodocyclis
gelatinosus: Possibility of horizontal gene transfer in purple
bacteria
Photosyn. Res. 36, 185-191 (1993)
2 (sites)
Nagashima, K.V., Matsuura, K., Ohyama, S. and Shimada, K.
Primary structure and transcription of genes encoding B870 and
photosynthetic reaction center apoproteins from Rubriivivax
gelatinosus
J. Biol. Chem. 269 (4), 2477-2484 (1994)
94132007
8300574
3 (sites)
Igarashi, N., Shimada, K., Matsuura, K. and Nagashima, K.V.
PHOTOSYNTHETIC GENE CLUSTER IN PURPLE BACTERIUM, RUBRIIVIVAX
GELATINOSUS
Photosynthesis. Mechanisms and Effects 4, 2889-2892 (1998)
4 (sites)
Menin, L., Yoshida, M., Jaquinod, M., Nagashima, K.V., Matsuura, K.,
Parot, P. and Vermeegh, A.
Dark aerobic growth conditions induce the synthesis of a high
molecular potential cytochrome c8 in the photosynthetic bacterium
Rubriivivax gelatinosus
Biochemistry (1999) In press
5 (sites)
Nagashima, K.V., Igarashi, N., Harada, J., Nagashima, S., Matsuura, K.
and Shimada, K.
Determination of Nucleotide Sequences of Rubriivivax gelatinosus
Photosynthetic genes
Unpublished
6 (bases 1 to 58996)
Nagashima, K.V., Igarashi, N., Harada, J., Nagashima, S., Matsuura, K.
and Shimada, K.
Direct Submission
Submitted (05-NOV-1999) Kenji VP Nagashima, Tokyo Metropolitan
University, Biology, Minamiosawa 1-1, Hachioji, Tokyo 192-0397,
Japan (E-mail:nagashima-kenji@cc.metro-u.ac.jp, Tel:81-426-77-2583,
Fax:81-426-77-2559)
Location/Qualifiers
1. 58996
/organism="Rubriivivax gelatinosus"
/strain="IL144"
/db_xref="taxon:28068"

FEATURES
source
CDS
gene
/clone="pgc#12, pgc#6"
/notes="pgc#12 (pos:1..34032), pgc#6 (pos:21943..58996)"
complement(128..1630)
/gene="pdx"
complement(128..1630)
/gene="pdx"
/codon_start=1
/citation=[3]
/gene="pdx"
evidence=not experimental
/transl_table=1
/product="exopolysphatase"
/protein_id="BA094021.1"
/db_xref="GI:7416772"
/translation="MADITTPELPLAAVDMGNSFRLIAIQRGRRSEYKDMWR
LGAGLDDQGLTEEAIVORGLCTRRPAATVGAPOVAATVQTLREARNRPAFLR
AENALGTEIVTSGREARLTIYAGVLOIOPADEPRIVYVIGRSTELIIGRGTPVRA
ESFPAVCUSLSIRPFQDGRIDAEAPRAAVAGAELEEMEPSSHHMOALGSSGTA
GAVSOILAAAGVSDGRITPEGLAWLEORCIAGHVRDLPGIKERRAVIPGGLAIL
QTIAENFGIATLAPAKGALRGQVIVDLHERLVALKTRPSRMDASVHELORFEDVY
KOARRVATVALAFQGVMPQADAEARELEGMACDLHEIGLMSHDSHSAVLAVY
DAPFSQSGQRRIQGLVIGRGGLRKIEBALASCELLGMQVCLRLAAIKCHARTVDVE
QALTLIRGRDAVLGIDATWAKAHRPTVTLSEAOQTMASGRKILLR"
1716..2591
/gene="ORF291"
1716..2591
/gene="ORF291"
/notes="similar to putative carboxymethylenebutenolide of
Methylobacterium extorquens AM1"
/citation=[3]
/codon_start=1
/transl_table=1
/protein_id="BA094022.1"
/db_xref="GI:7416773"
/translation="MQAEPDSLVSPDRNRDFTRTAVGSGFAAVLVPAQTLIT
DSAGLVGSEVITIPVDIFRPAAPANPAPLVVSEIFGVHAIIDVARLKA
GYFAVAPQLFVQGDAAATTEIPRLMSRVSVKVPDQVRDIDATVAMGQAGATYK
AGITFGCMGRITWYIAHSPOKGVAVRGLVGDKTALOPQVPELASQLAPYVG
LYGGDSGILPDSVERMKTALAAAGVAAQSFVYVPEAGHARNADYRSYADAAQ
GKRALAWFAHBGVG"
complement(2610..3104)
/gene="ORF164"
complement(2610..3104)
/notes="similar to orf1 protein (AAC05676.1) of Pseudomonas
aeruginosa"
/citation=[3]
/codon_start=1
/transl_table=1
/protein_id="BA094023.1"
/db_xref="GI:7416774"
/translation="MQVRSALVAPASHLPDLIEGAHYPAFLPCAGATTIVRDS
VVSADISIRKMGFEIRTRNPRRDEFATILIERPFRRFGEWCLKVLAEDACKYN
FLPDYEFDSAVTIRNAGPVDFDKIADLVAFQRAIDVPPVLAPAPAEAPAPAD
MPP"
complement(3134..4078)
/gene="ORF314"
complement(3134..4078)
/notes="similar to menA of Escherichia coli"
/citation=[3]
/codon_start=1
/transl_table=1
/protein_id="BA094024.1"
/db_xref="GI:7416775"
/translation="MEGERATVNSPPRAVTSMEAWAVRPTLPVAVSPVLVGLT
GDRICADIMAILVLTGAALMOVITNNQNDVGVVRGSESTGTRGLPRATANGML
SVRHRAGIVVALLAAAGVAVRGVPAIGASLALAAVYGGKPIAYTFEG
ELTVFVEGLIIVGTDVLTGSGGLVTLASVAGCAAAALAVNNHDDIANDRVG
RTPRAVTRGSSVALTYVLLIPRGELVPMALIAQSWTLPLALPSTIQLRPFV
HCPRGIAFVEIILFRTRLELNFPAALISAGALISMTAR"
complement(4158..4991)
/gene="ORF277"
complement(4158..4991)

```

```

/gene="ORF277"
/notes="similar to ubiquinone methyltransferase UblE"
/citation={3}
/codon_start=1
/translation={1}
/protein_id="BA94025.1"
/db_xref="GI:7416776"
/translation="MRRARGPVYKSSQPTLLDRLPRMNPPTLSRSQVSYDE
REERORIRPVSAVARYDMLNDLMSGIRLKRPRYRMAADPAQGHIDVLAGTSD
VALMAADRRVYVYDPSAMAVGQKGAHNDMDVGSKEQFLADASVDLITISG
IRNATRIDVALREIRHYLTKGRTLCLESTFPAWMLRPFNLVSFTVIFPLGMIAS
EAAITLVESIRKRDORGFAMITSAQFESVRWHDLSFGIACVHGTAAATPACT
A"
/gene=".6499
/gene="ORF468"
/5093. .6499
/gene="ORF468"
/notes="hypothetical enzyme for heme synthesis
(coproporphyrinogen III)"
/citation={3}
/codon_start=1
/translation={1}
/protein_id="BA94026.1"
/db_xref="GI:7416777"
/translation="MRTPTDHPDHPSPSVLKIANTLDAGASMCWQSPASPLVQ
GMLGNPSARLYEALRGADRPDEPLVSARVPCAAHCFCCDDIDPAODEYI
DVLVAGLVEIRVAVAGVLRGLLGLGGTASBELTGLARLVQAMQSOVRMDEYI
IDMSCECDPRVQOTMFEVLRGLGFRVSLVDLDVQOAGIRGOSALIDVCEI
ARAGIEYINLDLMTGLPROTMAWCAVTRGLIAMGPERILARVYRHPHARVAVI
DTSLSPALMAELARASAEMLNAGRWITGADQFVLEHDELSPRMDGRLRSITISY
TAAAPYPMIGLGVAGVEIDGAMFNWDSQAAMNLRHLHLPDSQARPATPSPVOR
AAVERILCTLELAADVGLDGVGRILAREAGLIVRVDRLVYVTEAGHNLALCT
DIUEPCAOPRGRARCC"
/gene="ORF247"
/6487. 7230
/gene="ORF247"
/6487. 7230
/citation={3}
/codon_start=1
/translation={1}
/product="hypothetical cytochrome with 4 heme binding
sites"
/protein_id="BA94027.1"
/db_xref="GI:7416778"
/translation="MRLNMRITATSPAAVGLAFALGPLALAPSDADAAAPAVM
APGAAPVAPASAPATHRPLRDLNLPQAPNPQSEPLFGYLDLPASFVVRKPHGLY
PCSQCHVPLINKOPRALVAAPVAVSLPHRGMMWICIDCHVADRLRTLDGRVDF
DOSOLVCGQCHGPRHRMAGAGKRAANNQGRQIYACTHCHDADHPKIPREPAFP
PVRAGLERPALTPAHAAQWPORAASGAR"
/gene="ORF321"
/7230. .8195
/gene="ORF321"
/7230. .8195
/citation={3}
/codon_start=1
/translation={1}
/product="hypothetical electron transfer protein with 4
Fes centers"
/protein_id="BA94028.1"
/db_xref="GI:7416779"
/translation="MSRQPSDPPDDPGQFNTPRRFLAGVATAGAVVTAATAGM
GGOTLDAMGSLTODHQRKTQDEIHAAALRIENNRRTGVDRCEDLPLRTVYFG
YALNISCKQKQRCQVCKENNLGRDSQOITRVLOMPGTDLEAHERYIDKQVY
VPHMTLPLVQCCQDNPPCYACPTKATMSEPGIIVVDWDCIGCKTCTACPTGAR
HFHWTFQILPAEFNPUNYLGRRPRKGVKCHFCOSARKGROPAQCEACTPTGAR
VFQNLDPKSEIKMVLNNVFRLEKEDLGEFKWYPSDDVSPA"
/gene=".9418
/gene="ORF408"
/8192. .9418
/gene="ORF408"
/8192. .9418
/citation={3}
/codon_start=1

```

```

Alignment Scores:
Pred. No.: 5,4e-07 Length: 58996
Score: 289.00 Matches: 94
Percent Similarity: 43.20% Conservative: 49
Best Local Similarity: 28.40% Mismatches: 143
Query Match: 17.31% Indels: 48
DB: 1 Gaps: 9

US-10-018-706-2 (1-322) x AB034704 (1-58996)
QY 10 GlnAsnGlnLysProIleValArgLeu-----GlyLeuIlePheGlyValIle 25
Db 53379 GAGAACTCCCGTCCCGACCCGCTTCCGCGGCGCTTCCGCTGCTGATGCGCGTG 53438
QY 26 ThrThrcysIleLeuAlaGlyCysAlaSerLysProThrcysAnserThrserySer 45
Db 53439 GCCCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53495
QY 46 GlySerHisArgThrserySerGlyGlyLeuAlaIleGlyserGlnValIle----- 63
Db 53496 CGCAGCGCGCGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53555
QY 64 -----ThrAspserGlnGlyValProAsnArgThrcysGlnValLysGlnGlyAsp 79
Db 53556 CCGCGCGCGCGCGAGAAACCTGGGCAAGCCCGGCTACTACCGCTCAAGCCCGGCGAC 53615
QY 80 ThrValserLysIleAlaGlnArgTyrglyLeuAsnTrpArgGlnIleGlyHisIleAsn 99
Db 53616 ACCGTATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCC 53675
QY 100 AsnLeuAsnSerSerTyrrThrlIeTyrrThrcylGlyTrpLeuThrcysLeuTyrrAsp 119
Db 53676 AACCTCAGAAACCCGACCGCATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCC 53734
QY 120 LeuIleValArgGlnArgSerIleSerSerGlyValAsnThrcysIleThrcysPro 139
Db 53735 TCCCGACCCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53793
QY 140 ValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThr 159
Db 53794 GTCCGG--CGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53835
QY 160 ProProValValValLysLysProThrProThrProProValValGlnGlnProAla 179
Db 53836 -----CCGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 53865
QY 180 ProValAlaPro-----ProValThrcylAlaProPheAlaThr 192
Db 53866 CCGCGACCGCGCACCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCGCGC 53924
QY 193 GlySerSerGlyValMetGlnPheArgTyrrProValGlyValAlaThrAsnProValAlaArg 212
Db 53925 GACGGGACGACGACGCTCAACCTGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53963
QY 213 ArgPheGlyThrcylAlaThrValAlaIleGlySerThrcylValThrseryAnserThrcys 232
Db 53964 -----GCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 54017
QY 233 GlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrcylIleGlnAlaPheHis 252
Db 54018 GGGCAGGCGCGTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 54077
QY 253 AsnMetAspGlyAlaSer-----IleValIleGlnHisThrseryPheValSerSer 270
Db 54078 GGCCTGCGCGCGTACGCGCAACCTATCATCTGATCAACGACACCAACCGCATCTTCTGCTGCGG 54137
QY 271 TyrlIleHisIleLysAspAlaGlnValIleThrcylAspThrcylValArgThrcylGlnArg 290
Db 54138 TACGCGCACCAACCAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 54197
QY 291 IleAlaSerMetLysAsnGlnProserGlyAlaAlaLeuPheGlnIupheArgIleSerArg 310
Db 54198 ATCGCGAGATGGCGCGCACGCGACGCGACGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 54257

```


AUTHORS	Kujat Choy,S.L., Meekins,D., Tindale,A.E. and Page,W.J.
TITLE	The stationary phase sigma factor (σ pos) of <i>Acetobacter vinelandii</i>
REFERENCE	Unpublished
REFERENCE	2 (bases 1 to 5876)
AUTHORS	Whelan,H.L. and Page,W.J.
TITLE	Aerotaxis in <i>Acetobacter vinelandii</i>
REFERENCE	Unpublished
REFERENCE	3 (bases 1 to 5876)
AUTHORS	Tindale,A.E., Meekins,D., Kujat Choy,S.L. and Page,W.J.
TITLE	Direct Submission
REFERENCE	Submitted (19-SEP-2001) Biological Sciences, University of Alberta,
JOURNAL	CM405 Biological Sciences Bldg., Edmonton, Alta T6G 2E9, Canada
REFERENCE	4 (bases 1 to 5876)
AUTHORS	Whelan,H.L., Kujat Choy,S.L., Meekins,D. and Page,W.J.
TITLE	Direct Submission
REFERENCE	Submitted (06-MAY-2002) Biological Sciences, University of Alberta,
JOURNAL	CM405 Biological Sciences Bldg., Edmonton, Alta T6G 2E9, Canada
REMARK	Sequence update by submitter
COMMENT	On May 6, 2002 this sequence version replaced gi:16226162.
FEATURES	Location/Qualifiers
SOURCE	1..5876

gene
/cd_1165="taxon:354"
/1..396
/gene="pcm"
<1..396
/gene="pcm"
/codon_start=1
/transl_table=11
/product="L-isopartate o-methyltransferase"
/protein_id="AAL16083.1"
/db_xref="GI:16226163"
/translation="LEIGTGGYQTATVAQVVERVYRQALQERAKRIELINR
NVEFMGMEKMFALPANGINVTANASDVFSALDQLAGRNVLIFVSGAEQQL
LIVREDDGFRHUALETVRVFLQGSLE"
1033..1686
/gene="mep"
1033..1686
/gene="mep"
/note="similar to nlpd gene of Pseudomonas aeruginosa"
/codon_start=1
/transl_table=11
/product="murcin endopeptidase"
/protein_id="AAL16090.1"
/db_xref="GI:16226164"
/translation="KRWGEMREHLAHHNGIAPVYIHFGKTIQUGPVYRQAASIAAP
SVPLATPVPYPAISQAAVAAPVAPDSGKVPPEPTATASPAQAVSRKSGSGMAYP
EGPLATRRSSNLSKNGIDIGGQADQPLASDSVYAGGLKRGVLEGLTIKSDYV
VSAVGHNRRLLVREQQVRAQVIAEMKSTGTDRVKFLIRKRGKVDPLQYIPSR"
1786..2790
/gene="tpos"
1786..2790
/gene="tpos"
/note="transcriptional regulator"
/codon_start=1
/transl_table=11
/product="stationary phase sigma factor"
/protein_id="AAL16091.1"
/db_xref="GI:16226165"
/translation="MALNKAKEAPFVDDVLIMDAIVLEVDLDEPDLPVAKRR
QNLQKHXYETRALDNTQYLINIGSPILTAEEVHFARLQGDGPARKKMTISSE
NLRIIVKALARYVNRGSLDLDLIEGNIGLIRAEKTPBGRFSTYATWIKQITTE
RAINQVQTRITRLPIHVKEILNVLYIAAEELTQKLDHSPSAEILNLEKRPVAEYKAKR
GIMBETVSDVSLGPDSDKTLVDLVTDRPBDPCLQDDDLQSIDMWLTDEKQR
EVVYVRFGIRGESCITLBEVQGEIGLTRERVRQIQVEALKRLREILKXNGISDALPQ
"

gene
3415..4356
/gene="aerp"
3415..4356
/gene="aerp"
/note="aerp"
/codon_start=1

gene
3415..4356
/gene="aerp"
3415..4356
/gene="aerp"
/note="aerp"
/codon_start=1

[illegible]

[illegible]

TITLE

JOURNAL
MEDLINE
PUBMED

REFERENCES
AUTHORS

2 (bases 1 to 10026)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvaranga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordini,S., Bove,J.M., Brites,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.B.A., Carraro,D.M.,
Carier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
Costa-Neto,C.M., Coutinho,L.B., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorry,H., Facinani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., France,S.C., Franco,M.C.,
Frohme,M., Furian,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hohnselt,J.D., Junqueira,M.L.,
Kemper,E.L., Kitzajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsumura,A.Y., Menck,C.F.M., Mitecka,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.G., Pereira
Jr.,H.A., Pequero,J.B., Quaggio,R.B., Roberto,P.A., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawaaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
Tshahko M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
Vettore,A.L., Zago,M.A., Zatz,M., Zeldin,J. and Setubal,J.C.

Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

FEATURES
Source

Location/Qualifiers

1..10026
/organism="Xylella fastidiosa 9a5c"
/db_xref="cazon:160492"
/clone="9a5c"
86..1393
/gene="XF0851"
86..1393
/gene="XF0851"
/note="similar to SP|P29011 (percent identity: 54 %/query
alignment coverage: 95.6 %/subject alignment coverage:
96.3 %) identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder/Start codon shift: -6"
/codon_start=1
/transl_table=1
/product="D-amino acid dehydrogenase subunit"
/protein_id="A83661.1"
/db_xref="gi:9105759"
/translation="MRVLIIIGSGYIGTTSAMYLSKAGEVVDVROSAGLETSTAN
GQSLFGTSPAPAGVPEPLKAVKMLPERPAALISIPPTDMDVYVLAARLHCSARVY
VNKSRMLRSLSEYSEALASLAEFTGITEEGHGLCTIQLEFQQOQLDVAVRIELTLO
GIPEVSLSPQISKEFPGIADGSVAPFQALPHDQDCCIFQRLALAAKRAVE
RYCGTVRLVDGPRVIGALNGLERADCCVVALGSYPLLAIPGLRFLPVLPKAVE
SLTLPMDASRAPVSTVLDSEYKAVIRFDRIRVAGAAVSGVDSINPRLATLE
VVDVYRGGDIDGGERFTGCRPATPPTCTPIVCAVPGGLFINTGHGTLWTMSGGS
RYLADLICRPCEISSEGLDMFRVLVTIPQEQACAPCPPTP"
1363..2460
/gene="XF0852"
1363..2460
/gene="XF0852"
/note="similar to SP|P06655 (percent identity: 45 %/query
alignment coverage: 98.4 %/subject alignment coverage:


```

100.0 %; identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="alanine racemase"
/protein_id="AAF83662.1"
/db_xref="GI:9105760"
/translation="MCSVRPAHVIDLDLRHNYRFRRLGGKALAVKADYHGA
VCAQALAEVDGFAVACIEEALRQAOIAPILLGFEAEIELELAKNLMVY
ASSQVRLAAAFHSPSPIGVWLKLDGMRHLGLEPFRDAMSLHRLQVEIVMT
HARADELNPRTEQASVPAQSQGVAPASQSSGLGWAQYSDMRQGLMYG
VNPIDPSTSLGSLPRVMMVYSRLIAVHELPAQEPVYGACFVTERPRVGVAMYG
ADGYPYAVNGTFLVDGVCPLIGRVSMDMLTVDLTDHPQADVQVQVQMGVQDVA
ELAAHNLAYQLCGLAKRVRYYLGEMGAV"
2937. .4436
/gene="XF0853"
2937. .4436
/alignment="similar to GI|2765035 (percent identity: 37 %/query
alignment coverage: 75.8 %/subject alignment coverage:
50.8 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="two-component system, sensor protein"
/protein_id="AAF83663.1"
/db_xref="GI:9105761"
/translation="MOTVLPDCAVDGIPRLTNASFGPHLRIGSVVGEGLCDIMPDG
PQIRKRGVALSCGRHREPARLCEWYSTVWVWVNGAPRGRLGVLPMVTEKRMAR
KLAESEKRVYLDAGLPYVVDAGCARPVNTATGAPGNDLSSGTGEMVWAP
EDVLEIFKLSALKTQSFRLVYARARYCOMRIFMSAPPSDGHGICITGSSP
DVTRELELARKQLKSEKRAENENARLQDBELTSHLPRPLTILICMSBL
LVRIEDHPCIKGLSYLASAKYQKRLISMDLSDLSLCKMLKKEVLDIVCEIET
LGIYEPVDSKQRLRLAREMPCVLVDKTRKQI FENLSNAIKYTPDCCIDPEI
DAGDSFRSVTSDGDLAEPLPHLSRFRQADGTTTCHGELGGLAVQSLVEMH
GGYVGAISFGIGAMFTVSLRYVCAIDVADVLICRRIGSAGDQCY"
4712. .4921
/gene="XF0854"
4712. .4921
/alignment="similar to SP|P45682 (percent identity: 39 %/query
alignment coverage: 112.8 %/subject alignment coverage:
91.9 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder/Start codon shift: -144"
/codon_start=1
/transl_table=11
/product="lipoprotein"
/protein_id="AAF83665.1"
/db_xref="GI:9105763"
/translation="MERSEMSKLLSTALALATACSTATTVPNNSTNSRTAKEN
QTVVKGDDTLAISRTGVAPODIAANRNLTASKTIYGGVRLYEDASAPSTPQ
FTVSRPTSPNSPTTITAPANSQGNWNPTEGAVVSVFVAGQTTKQGVASINGNGQT
IRAAANGTVVSGALIGYELLIIKHBEOWISAYGHNKRLVNEGQTVKANQDIAEM
GRMYLFEIRYNGKVPDPLVLPKK"

```

```

gene      complement (6327. .6944)
CDS       /gene="XF0856"
          complement (6327. .6944)
          /gene="XF0856"
          /note="similar to SP|P36684 (percent identity: 37 %/query
          alignment coverage: 97.6 %/subject alignment coverage:
          104.7 %); identified by sequence similarity; putative; ORF
          located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AAF83666.1"
          /db_xref="GI:9105764"
          /translation="MKTFGLYRNRAISNRHVPALTLGSPVEGFPPVPEVMT
          LPSIANRHQALWFAVLSLGSALGILGHRAPVMDPIAMIGMSAIDKQIH
          QLOQLVSPWRAFWLVNGFTPIPLKFTVAGSVGVPLPFLSSMLIGRKYVL
          VAGAIKIGKRAETLHRMIEPLGITSLALIGLVWVWKTKFG"
          complement (6964. .7641)
          /gene="XF0857"
          complement (6964. .7641)
          /gene="XF0857"
          /note="similar to SP|P24206 (percent identity: 47 %/query
          alignment coverage: 93.3 %/subject alignment coverage:
          101.0 %); identified by sequence similarity; putative; ORF
          located using Glimmer/RBSfinder/Start codon shift: 39"
          /codon_start=1
          /transl_table=11
          /product="L-isoaspartate O-methyltransferase"
          /protein_id="AAF83667.1"
          /db_xref="GI:9105765"
          /translation="MTAPSLQAKVAGIGMTSQRVDRIVERLRECGIODEKVLTTIR
          IYPRHLFIDEALRAYEDTALPIHGQTI SQPVVAMTAVMWGVAPEKILGIGS
          GYOSALIASGELAVTIERIGLLOARRRFROLKIRSKDDSGISGTEHPANAI
          LVTAAPADITDLILQALGRLVAVPGVASQALVOLRTIDGNITHEILPVEVS
          ILRQVLD"
          complement (7638. .8426)
          /gene="XF0858"

Alignment Scores:
Pred. No.:      1,036-05      Length:      10026
Score:          255.00      Matches:      87
Percent Similarity: 39.02%      Conservative: 32
Best Local Similarity: 28.52%      Mismatches: 168
Query Match:    15.15%      Indels:      78
DB:              1          Gaps:      9

US-10-018-706-2 (1-322) x AE003925 (1-10026)
Qy      20  LeuilePhecllyVallleThrThrCysileLeuAlaGlyCysAlaSerlySPothrTyR 39
Db      6300  CTATTACTGTCACACAGCGCTCGACCTAGCACTGCTGCGACCTGCGACTGCGACTGCGA 6241
Qy      40  AsnSerThrSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGly 59
Db      6240  CGCCCAACAAATACGACCAATCACCACACCGCC----- 6205
Qy      60  SerGlnValIleThrAspSerGlnGlyValProAsnArg---TyGlnVallyGlnGly 78
Db      6204  -----AAACCAACCAAACTGCTGTTGGAACAGGCC 6172
Qy      79  AspThrValSerlyIleAlaGlnArgTyGlyLeuAsnTrpArgGlyIleGlyHisIle 98
Db      6171  GACACACTGATAGCATCTCGCCCGGACCGGCGTTGACCGCAAGATCTCGCCGATGG 6112
Qy      99  AspAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeuThrLeuTrpSerGly 118
Db      6111  AACGAGCTCAGCTGATCAACAAACATCTACCGCGGCAAGTACTGGCTTGTATCCAGAG 6052
Qy      119  AspLeuLysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSer 138
Db      6051  GAT----- 6049
Qy      139  ProValAlaValGlnSerSerArgProValAlaGlnGlnHisProAlaValAlaGlnLysPro 158

```



```
/gene="PM1607"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAK03691.1"
/db_xref="GI:12722004"
/translaction="MRLPITFLVAVLLFQYDFEKGNGLYDKRTAQQAQKQENR
KLSQRNOVAALIEDLKQGVAEERARFQHDVWKEPDEIYHIVKEOK"
5163. .5879
/gene="PM1608"
/gene="PM1608"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAK03692.1"
/db_xref="GI:12722005"
/translaction="MTTPIAKSTRRIYAVVPAAGISGRMQMDKPKOYLHIGKTLIEH
TLVILGYPLIEKIILAVANDPYISTCPILITRKIQLVESGSSRADSVLNGNAVNS
AVQNSDEPMVWHDARPCLTODLKVQVEDNGAIIAIPADITIKRALHNOQIHY
TEDRSQMLAQTPOFPPIATLQALFQALQGLQVTDASAMEFAGFRPHLVAGRSIN
IKVTRPDPFALAEYLSRTK"
5888. .6364
/gene="PM1609"
/gene="PM1609"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAK03693.1"
/db_xref="GI:12722006"
/translaction="MIRIGRFPVHAHAGSGPILIGGVAIPYEKGLAHSGDVATLA
LTDALHVALAGDVGKLPDPTDQKQADSGRLERAYTOAGYGVGVQVDTPTIO
AFKRPFLDARAKAIAEDLACDIEQVNVKATTSRLGTNGEGIAERAVLVAKS"
6364. .7371
/gene="PM1610"
/gene="PM1610"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAK03694.1"
/db_xref="GI:12722007"
/translaction="WELAYLHTREPOVARLKAECADFIKENLGYEMSGDEFAVAVY
KRTDCNTLTFVGEKLAQFVGISERNNGVAGLDRKAVTEOMFCLMPOGPTDFPSFOL
EGVDILEVTRHRRKIRITSGLEGNHEIILRGAKETDELNVLNHIKQCGFPNFTBOR
FGRDGHNTQALRWAGSEINVDKRRKSFYLSARSEVFNIVSERIALQIAQOVLHG
DMLQOGSHSFQADEKEDNALQRLREOQDILITAPLIGQNPATDIEQLVEORQ
ALITMAKERKKAARRPMLQOALQWAFVAGKLAAYFLPAGSVATFALVREVVQDE
D"
7401. .7919
/gene="PM1611"
7401. .7919
/gene="PM1611"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAK03695.1"
/db_xref="GI:12722008"
/translaction="WORMSKTILSLMLVSTNNVFAQEEEDPVFEGQVYAEAKRDHAT
AFKMLSRAGSGEMSAQFNVGKMTDDGGVRODQKQALKTYQKSHQNHDPACQTHLGI
MYSBGDGLADPFQKAYKWSQSAVQDARALYNIGTLVANGEGIERMDRAKMTFKQA
CKAGLPEGCDWE"
7991. .8731
/gene="PM1612"
7991. .8731
/gene="PM1612"
/codon_start=1
/transl_table=1
/product="Surg"
/protein_id="AAK03696.1"
/db_xref="GI:12722009"
```

```
/translaction="MKILLNSDGIHAEGIQILAREKRFADVTLVAPDRNRSAASS
LITLPERLRLRPNQDYGCLNGPADCVLALNGLSGVDLVGSGINGVINDGVITY
SGTVAALAEGRYLGPAIYAVSDGROHYESARVVCIELPRHQIILQORELINTNPV
DIPEYEEKGVKCHGYRAAAAEVVKODPREEALYVMQAPALANEQEGTDFHAKN
GVYATPIQADMTAVHSLQSLQDWLSE"
8771. .9346
/gene="lppb"
/Note="PM1613"
8771. .9346
/gene="lppb"
/codon_start=1
/transl_table=1
/product="lppb"
/protein_id="AAK03697.1"
/db_xref="GI:12722010"
/translaction="MKIRFAMVDKTMANKRHYATFRFALTSFIAIPRPIDMYLI
PMSISOPRALKAFITAVASVGVYIGLGLTAVDVQVHYQOPGTOCHETISM
FERKGLIVFVAGFSEIPYKVFITCAGVQWMLFPLPVLATISRARFLVAKLAWG
GEKPAKLRSLIELIGMSVLLAVIAYLILR"
9559. .10962
/gene="PM1614"
9559. .10962
CDS
gene
CDS
Alignment Scores:
Pred. No.: 2,21e-05 Length: 12956
Score: 249.50 Matches: 80
Percent Similarity: 45.69% Conservative: 42
Best Local Similarity: 29.96% Mismatches: 110
Query Match: 14.94% Indels: 35
DB: 1 Gaps: 9
US-10-018-706-2 (1-322) x AE006198 (1-12956)
Qy 73 TyglnValIsgInglngIysprrThrValIserIyIleAlaIglArygrIyLeuaenTrp 92
Db 10201 TATACGCTGCGTAAAGCCATTAACATGACCTATGCTCATATTCACGGCTTAGATGTC 10260
Qy 93 ArgGluIleGlyHisIleAsnAsnLeuaenSerSerTrpThrIleTrpGlyGlnTrp 112
Db 10261 AAGAAATGGCATCATTAACATATGCTCGAACCTTACCTTAAGCTGAGGCAAAAC 10320
Qy 113 LeuThrIleutpserGlyAspIleuIyValArgIluArgserIleSerSerGlyValaen 132
Db 10321 TTACGTGTGCGATGGA-----CGTGGCAAGCACCTCTTCACCAACAGTG 10368
Qy 133 ThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152
Db 10369 ACA-----CAACCTGAGACGGTACCCGCTTCACMACCG-----AAGACCACT 10410
Qy 153 ProAlaValGlnIysProThrProPro-----Val 162
Db 10411 GAAGTAACTATACCCCGCGGTCACATGTACCAATACCGTTGCGATGTGATCAATCAT 10470
Qy 163 ValValValIysIysProThrProThrProProValValGlnIlnProAlaProValAla 182
Db 10471 GAGCAATTAATACAGGTGCAGTCCGCCCTGTTCCAGTCCAACTGAACCAAGTGTT 10530
Qy 183 ProProVal-----ThrguAlaProPheAlaThrcIySerSerGlyValImet----- 198
Db 10531 AAGCTTTTAAAGCAAGCTGTGATACCTGTCGCAAGCAGGAGTAAACATAGTTTGG 10590
Qy 199 -----GlnPheArgTrpProValGlyAlaThraAsnProValValArgAsgPheGlyThr 216
Db 10591 AATGTAACTGGCAATGCCAACCAAGCAAT-----ATTTCACAGGCTTTTCAAC 10644
Qy 217 AlaThrValAlaGlySerThrValItnrSerAsnGlyIyMetTrpPheSerGlyAsgAsgGly 236
Db 10645 GCAGACGGTGCC-----AATAAAGGATTAATATTCCTGCTCAGCTGCT 10689
Qy 237 AspIeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAsgGly 256
Db 10690 CAAGCGGTGAATGCCGCCGCGGACGAGAGTGTTATATCGGGGTAAATGATGATGCTGT 10749
```

QY 257 Alaser-----llevallelleglnhiethrasnglyphevalserSeryllehielle 274
Db 10750 TATGGCACTTAATTATCATTAACACATGATCATTAATTAAAGCTTAAGCTCAAT 10809
QY 275 lyahspalaglnvallyethrglyaspthrvalargthrglyglnarglllealaserMet 294
Db 10810 GAAAGATTATTAGCAAGATCAAGCAAGATTGTCGGGTCAACAGATTGCCAATG 10869
QY 295 lysannglnproserglyalaaleupheglupheargleserarganlyvaltyr 314
Db 10870 GGTACCTCTGTGACCATAGTCAACCTTCAATTGAAATTGGCTTAAGGCAATCA 10929
QY 315 ValaspProleuthrValleu 321
Db 10930 GTTCATCCACCGGTATTG 10950
RESULT 11
PPY19122
LOCUS
DEFINITION
Pseudomonas putida nlpd (partial), rpos, fdxa, muts (partial)
genes.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS
gene
CDS

PPY19122
Pseudomonas putida nlpd (partial), rpos, fdxa, muts (partial)
genes.
Y19122.1 GI:5305138
fdxa gene; ferredoxin; lipoprotein; mismatch binding protein; muts
gene; nlpd gene; RNA polymerase sigma factor; rpos gene.
Pseudomonas putida.
Pseudomonas putida.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 3539)
Kojic,M., Degraessl,G. and Venturi,V.
Cloning and characterisation of the rpos gene from plant
siderophore and homoserine lactone production
Biotechm. Biohyb. Acta 1489 (2-3), 413-420 (1999)
2 (bases 1 to 3539)
Venturi,V.
Direct Submission
Submitted (24-JUN-1999) V. Venturi, International Centre for
Genetic Engineering & Biotechnology, Bacteriology Group, Patriciano
99, 34012 Trieste, ITALY
location/Qualifiers
1..3539
/organism="Pseudomonas putida"
/strain="WCS358"
/db_xref="taxon:303"
1..735
/gene="nlpd"
<1..735
/gene="nlpd"
/codon_start=1
/transl_table=11
/product="lipoprotein"
/protein_id="CAB46190.1"
/db_xref="GI:5305138"
/db_xref="SPTREMBL:O9MWV6"
/translation="KLLVLLALWGTLLAGSSSTSGTSARVVDNRNNAAPKRPVTSQGY
IVKGDYLFISIAFYGMDYKELARNGI PAPIYIRPQPIRFSGSGTSTVVSPPS
SKRTVIRRPVSGSGASPASTKATPAPIPAVAVATVPAEAAVAGMTMPANGVLI
GKPSVNSGLKGIADIAGLGPVPAASGAVVVDGSLRGVAILIIKHSDTVYSAVG
HNRLIYRBOQVYKAGSIAMVHGH"
920..1927
/gene="rpos"
920..1927
/gene="rpos"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="CAB46191.1"
/db_xref="GI:5305130"
/db_xref="SPTREMBL:O9MWV5"

/translation="MALNKAPEFDIDDDVLLMETGIVLETVDSDEPAVPSVTRAK
SGSSLKQHYIDYTRALDADTOLYLINEIGSPILSPSEEVAFALSGQKPAGRKMI
SNRLVVKIARVYVNRGLDLIDIEGNGLIARVKEPDERGFRSTYATWIRORI
EPAINNOTRTIRPIHVKVKNVYLAARLQKLDHPSPERIATLEKPAVEVM
LGINERVSVDVSLSPDSKTLDTITDPRPTPCELIDDDISGSIIDQGLTDRKO
REVVVRRFGIRGHSSSTLDEVGLGILTRERVAQIOVEGLKRLREIKENGLSSSLF
0"
complement (2429..2752)
/gene="fdxa"
complement (2429..2752)
/gene="fdxa"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="CAB46192.1"
/db_xref="GI:5305131"
/translation="MTFVVTDNCKIKKTYTDCVEVCVDCYEGSPNLFVHPDECIDCA
LCBPCEPQAIFFSEDEVPSGMENFISLNLMLAIWPNITERKDALPDAERWDGKTGI
ADLER"
complement (2873..3539)
/gene="muts"
complement (2873..3539)
/gene="muts"
/codon_start=2
/transl_table=11
/product="mismatch binding protein"
/protein_id="CAB46193.1"
/db_xref="GI:5305132"
/db_xref="SPTREMBL:O9MWV4"
/translation="SFVPAQCELSLVDRIFTRIGSSDDLGRSTFVMESETPANIL
HNAITDLSLIMDEVRGTSTPDGLSLMAAABEALQRLAYTATYFELTYLPESBP
LVANVNLNTEHNERIVPLHNVLPGRASOYGLAVOLGVPFAVIORAEHLGRLET
ASLPHQPHSTAKDALHVRHQSDFASLPHRAIETGLDLDMDTPROAIENLYOLK
NLL"
BASE COUNT 678 a 1006 c 1098 g 757 t
ORIGIN
Alignment Scores:
Pred. No.: 9,74e-06 length: 3539
Score: 245.00 Matches: 95
Percent Similarity: 43.59% Conservative: 41
Best Local Similarity: 30.45% Mismatches: 119
Query Match: 14.67% Indels: 57
DB: 1 Gaps: 14
US-10-018-706-2 (1-322) x PPY19122 (1-3539)
QY 20 leuilepheglvallelthrthrcysileleualaglycysalaserlysprothrtyr 39
Db 7 CTGGTGCTTGCACTGGACATGGACACTTGCTGCGGGTTC----- 48
QY 40 AmsSerThserglyserglyserhlaargthserglyserglyleualallegly 59
Db 49 ---TCCAGCACACAGCTGCAGACGTGGCGGTGTGTGACCCGTAAACAATGACGGCCCAAG 105
QY 60 SerGlnVallelthrapserglnnglyvalproamnarhgrlyrlnalllysglnlyasp 79
Db 106 CGGCCGACGGGTGACCTCCGGGCA-----TACATCGCAAGCCGGGGGAT 150
QY 80 ThlValserlylellaaglnarglyrglyleuanstrparaglyluilegylhialeasn 99
Db 151 ACCGTGTCCTCAGCTCCGCTTCAGCTGGAGACTCAAGAGAGTTGCGTCCCGCAAC 210
QY 100 AmlleuanserSeryThriletyrthrglyglntrpleuthrleutpserglyasp 119
Db 211 GGCATCCCGCGCTTAACCATTCGCCCGCGCAGCCGATTCAGACAGCGGT--- 267
QY 120 leuilevalarglnargserileserSerglyValasnthr---AlaHISThrproSer 138
Db 268 -----TCCACGGGAGACACACGGGTGTCACGCCGTG 303
QY 139 ProValalaValGlnserSerArgproProValGlnGlnhiisProAlaValGlnlyspro 158


```

protein_bind
/1391..1408
/note="central position to predicted promoter: -208"
bound_molecety="Phob predicted site"
1575..1602
/note="factor Sigma70; predicted +1 start at 2990069"
1656..3032
/gene="ygeH"
/note="b2852"
1656..3032
/gene="ygeH"
/function="putative factor; Not classified"
/note="0458; This 458 aa ORF is 28 pct identical (6 gaps)
to 407 residues of an approx. 560 aa protein IAGA_SALTI
SW: P43016"
/codon_start=1
/transl_table=1
/product="putative invasion protein"
/protein_id="AAC75891.1"
/db_xref="GI:1789216"
/translaction="MDLENKFSYHLEGLTLEDGILTOGNEQYVYPOKEGLVILV
ESAGHVLKMLISWKNIIIVSDSEITRCYISRCIFPKIGYDRCIETIYRKYRPS
GQVKTINENNTSDYSIAIFPPTTSANTDPLINDELVOIINNKIDGITYPMMA
TNPNDHISQNSPLSRFPDYFVGTGRINONNAVTLITELIDAKNLFLINSHLPVE
LHNTSOFIINDILOTVAHPEKSVRLAKODQYKNHYISDEMLAKKELVPTPESYIR
AMFIDRLQNSDIOITKTECYCLACHNSLHAGSELELAOKKLELDYSDIT
TVDGKILMGLITGLSGOAKVSHILPEOAKIHSTIDASLYYVALVHPNKEIEAR
ICIDKSIQLEPRRRKAVIVKECVDMYVNPKNKIKLYYKTESSESRVVIDNLTKK
QUTRICR"
3039..3061
/note="central position to predicted promoter: -100.5"
bound_molecety="Fur predicted site"
3039..3061
/note="central position to predicted promoter: -113.5"
bound_molecety="Fur predicted site"
3118..3145
/note="factor Sigma70; predicted +1 start at 2991612"
3130..3158
/note="factor Sigma70; predicted +1 start at 2991625"
3200..3418
/gene="b2853"
3200..3418
/gene="b2853"
/function="orf; Unknown"
/note="072; This 72 aa ORF is 33 pct identical (1 gap)
to54 residues of an approx. 2520 aa protein TUD_DROME SW:
P25823"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/protein_id="AAC75892.1"
/db_xref="GI:1789217"
/translaction="WTFNIGINISGSNNIATGDEVSLDKHINTSATDIOQAF
IVSTWAPFQNDMYSBPNISPYKIE"
3521..3549
/note="factor Sigma70; predicted +1 start at 2992016"
3561..3977
/gene="b2854"
3561..3977
/gene="b2854"
/function="orf; Unknown"
/note="0138; This 138 aa ORF is 39 pct identical (7 gaps)
to 128 residues of an approx. 168 aa protein IAGB_SALTI
SW: P43018"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/protein_id="AAC75893.1"

```

```

/db_xref="GI:1789218"
/translaction="MSFASNSCSNAGTWERIPNLIKAYALVESNLKQDSIGNRDK
NNIKISIDYLMQMIHPIPLKRGKIDKIDRILNPNCLINIKIGIEILVNHRSRCGV
TWICITGVNAGFPMQNDQKROOYAPKILVILPOLMN"
complement (4022..4468)
/gene="ygeK"
/note="b2855"
complement (4022..4468)
/gene="ygeK"
/function="putative regulator; Not classified"
/note="E148"
/codon_start=1
/transl_table=1
/product="putative 2-component transcriptional regulator"
/protein_id="AAC75894.1"
/db_xref="GI:1789219"
/translaction="WKGAEIVKVKSHKIDAHITTVAKKPYIDSKLEBAKGCW
KTHPAPKLNADISNGYTPDPSVHNDCKISSRSDQITNRESEILQIADGKT
NKEIAPFLQLSRKTVEYTHRLIMKIDVHSGIELIKITLMMGCTI"
complement (4499..4654)
/gene="b2856"

```

gene

Alignment Scores:

Pred. No.:	7,71e-05	Length:	9720
Score:	238.00	Matches:	78
Percent Similarity:	39.62%	Conservative:	46
Best local Similarity:	24.92%	Mismatches:	103
Query Match:	14.25%	Indels:	86
DB:	1	Gaps:	8

US-10-018-706-2 (1-322) x AE000369 (1-9720)

```

QY 16 LysArgLeuGlyLeuIlePheGlyValIleThrThrcysIleLeuAlaGlyCysAlaSer 35
Db 9429 AATCTCTGGATTCGATGATGTTGTCGTTGAGCTCTTTGGGGGCTG----- 9376
QY 36 LysProThrTyraSerSerThrSerGlySerGlySerHisArgThrSerGlyGly 55
Db 9375 -----TCGGGTGCAATCATCCGATCAGAGACGATTCGGGC 9337
QY 56 LeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrglnVal 75
Db 9336 -----TCGGTTTACACCGTG 9322
QY 76 LysGlnGlyAspThrValSerIleAlaGlnArgTyrglyLeuAsnTpaGlnIle 95
Db 9321 AAACGGGGGATTCGCTATATCGATTTCGGCACACCGGAAACCAACGTAAGAACTG 9262
QY 96 GlyHisIleAsnLeuAsnLeuAsnSerTyThrIleTyThrGlyGlnTyrPleuThrLeu 115
Db 9261 GCGGACTGAACGGCATTTCCCTCCCTTACACCATTAAGTTGTCGAAACTTAAACTG 9202
QY 116 TrpSerGlyAspLeuIys-----ValArgIubArgSerIleSerGlyVal 131
Db 9201 -----GCTGGGGGAAAGATAGACGATTAACGTAATCAACCGCAATCAACGACC 9148
QY 132 AsnThrIleHis---ThrProSerProValAlaValGlnSerSerArgProProValGln 150
Db 9147 AAACCGCATCGGTTTACACCGTCTACCGCGGTAACGAAATCATCTCGCCGCGTGGG 9088
QY 151 GlnHisProAlaValGlnIysProThrProProValAlaValIysIysProThrPro 170
Db 9087 CAACGTTGTTGTTATGCGCAACGACGAGAAAGTTATCATG----- 9046
QY 171 ThrProProValValGlnGlnProAlaProValAlaProValThrGlnAlaProPhe 190
Db 9045 -----CCGTAT 9040
QY 191 AlaThrGlySerSerGlyValMetGlnPheArgTyrrProValGlyAlaThrAsnProVal 210
Db 9039 TCGACGACGAGATGGCGG----- 9022
QY 211 ValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrp 230

```

```

Db 9021 -----:~::~:
231 PheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla 250
Db 9006 ATCTCAGCTCCACCGCGGTACACCTATTATACCCCGCGGTGAGGAAGGTGTATGTC 8947
251 AspHisAsnMetAspGlyAlaSer-----IleValIleGlnIleThrAsnGlyPheVal 268
Db 8946 GGCACACCGCTGCTGCTGCTACGTAATCTCATGTGTAATAATGCGCAAGCGTGAAGCTGCGG 8887
269 SerSerTyrIleHisIleLeuAspAlaGlnValIleThrGlyAspThrValArgThrGly 288
Db 8886 ACGGCTTACGCCCATTAATGACACGATGCTGTAATAATGCGCAAGCGTGAAGCTGCGG 8827
289 GlnArgIleAlaSerMetLeuAsnGlnProSerGlyAlaAlaLeuPheGlnPheArgIle 308
Db 8826 CAAATAATCGCATCATGCGGAGACGCGGATGCGGATGCTGCGCTGCTCATTCAGAT 8767
309 SerArgAsnGlyValTyrValAspProLeuThrValLeu 321
Db 8766 CATTACCTGCACACGCAATTGATCTCGCTACGTTACTTG 8728

RESULT 13
LOCUS EC028375 55175 bp DNA linear BCT 08-DEC-1995
DEFINITION Escherichia coli K-12 genome; approximately 64 to 65 minutes.
ACCESSION U28375
VERSION U28375.1 GI:887800
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 55175)
REFERENCE
AUTHORS Plunkett, G.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-1995) Guy Plunkett, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined as part of the E. coli Genome Project
(Frederick R. Blattner, director) at the University of
Wisconsin-Madison. Supported by award HG00301 from the NIH Human
Genome Project. The entire sequence was independently determined
from E. coli MG1655; overlaps with other sequence determinations
are annotated. This entry should be considered somewhat
provisional; it will be updated and merged with others at a later
date.
FEATURES
source location/Qualifiers
1..55175
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
/map="approximately 64 to 65 minutes"
/notes="this sequence comprises the following lambda
clones: DD644 (EC30K464-1), DD645 (EC27-1154)
DD649 (EC22-159), and DD653 (EC18-341); M13Janus vector was
used for subcloning"
370..816
/notes="ORF_0148"
/codon_start=1
/transl_table=11
/label="ORF_0148"
/protein_id="AA083032.1"
/db_xref="GI:887801"
/translation="MKPRNINNSLPQPLVPDQENKKNKNEKSNVPIKMGSGLYN
IEQESLGKGLTIDLSIKIADISEETIQAIISAMSIYFSTIDDISAAVVELIKLT
KIENVNLIKFTTICTDVSFRTVTRHNKRNKHSYTLKNPLFNH"
1084..1575
/notes="ORF_0163"
/codon_start=1
/transl_table=11
/label="ORF_0163"
/protein_id="AA083033.1"

CDS
1910..3286
/notes="ORF_0458"
/codon_start=1
/transl_table=11
/label="ORF_0458"
/protein_id="AA083034.1"
/db_xref="GI:887803"
/translation="MDLENKPSYHLEGLTLEDDGILTQGNROYIPOKEGLVILV
ESAGHYVLMWILBESWKNIIVSDSLPRCIYSLRCIEKIGYDRCIETIYKGRFS
GOVFKYKIMEDNTSDISLIPFTSLNITDDELITNOELVQIISKKIDGLTYPMMA
LNCNDHISONSFLSRFPDITVGTGKINONNAVTLTXETIDAKXLFLIASNHLVDE
ANTDSQFIIDNLIQVHKRERSVRVLAQDQGYKXNYISDPMAGKGLVDPRESIYR
AMTFDLRKNSDIOTLKTECYCLAECHMSLAHGKSELALOKALELIVYSDIT
YVQKLIAMGILTGLSGAKVSHLFBQAKIHSITDIALITYRALVYHREKTEAR
ICIDKSLQLEPERRRAVVIKECVDMVNPVPLKNNIKLYKTESESHRYIIDNIIKTK
OLTRICMR"
3454..3672
/notes="ORF_072"
/codon_start=1
/transl_table=11
/label="ORF_072"
/protein_id="AA083035.1"
/db_xref="GI:887804"
/translation="MTNPIGINNLSQSSNIANATGDEVVSLDKHTINTATDIOIQA
IVSTWMAFQNDWYSEDNPISPYKIEW"
3816..4232
/notes="ORF_0138"
/codon_start=1
/transl_table=11
/label="ORF_0138"
/protein_id="AA083036.1"
/db_xref="GI:887805"
/translation="MSFANSCSNEAGTWRIEPLNLIKALVESNLKKDIOGNRX
NNNRIKSDWLMQINOMHIFILKRGIIIDERDLNPLCNIKIGTEILYNHRSRGV
TWOCIGYVNGPMDNCKRGOVAPRYIILYELMAN"
4277..4723
/notes="ORF_0148"
/codon_start=1
/transl_table=11
/label="ORF_0148"
/protein_id="AA083037.1"
/db_xref="GI:887806"
/translation="WKGAEIYKWKVSKHKIDAHITPRVAKPYIDSTIKLEAGAKGW
KTSHPARKLNPAIDISNGTYTPSDVHDDCKISRSDDQLNRESILQIADGKT
NKEIANFLQSRKIVETHTRLNIRKLDVHSGILEIKALNAGCTI"
4909
/notes="ORF_051"
/codon_start=1
/transl_table=11
/label="ORF_051"
/protein_id="AA083038.1"
/db_xref="GI:887807"
/translation="MGKIKIVSDQCFMIDGIIIGLGHVDDLYEVVGYKDKKALA
ECKNSTA"
5131..5562
/notes="ORF_0143"
/codon_start=1
/transl_table=11
/label="ORF_0143"
/protein_id="AA083039.1"
/db_xref="GI:887808"
/translation="MINDIKSIIILKPEEVDVFIKIPESWTKLPSISGPNLHIPS
FKDSIRVESYFVDSKIMNVIITPHDCKRFVFPDQFIAESVPEFVDCNGCOYIINN
CSPDKVNEICEORAHYIYKMEFETHLDMNNSVLAAPEDL"
5565..5837
/notes="ORF_050"
/codon_start=1
/transl_table=11

```


Oy		289	Gtargtgleaaasermetylaasglnproserglytaalaalephglhphearglle	308
Dn		9081	CAAAATTCGCCCATCTGGAGACGCGATCGGCCTCATTCCAGATT	9022
Oy		309	SearagysnglyValTyValAspProeuhrvalleeu	321
Dn		9021	CCTTACCGTGACAAGCAATTGATCGGTACTGTTACTTG	8983
RESULT 14				
LOCUS	AP002563/c			
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 14/20.			
ACCESSION	AP002563			
VERSION	AP002563.1			
KEYWORDS	GI:13363121			
SOURCE	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)			
ORGANISM	DNA.			
REFERENCE	Bacteriophage phi O157:H7			
AUTHORS	Escherichia.			
TITLE	1 (sites)			
JOURNAL MEDLINE REFERENCE	Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasekawa,C. and Shinagawa,H.			
AUTHORS	Complete nucleotide sequence of the prophage VT-2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak			
TITLE	Genes Genet. Syst. 74 (5), 227-239 (1999)			
JOURNAL MEDLINE REFERENCE	2 (sites)			
AUTHORS	Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.			
TITLE	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655			
JOURNAL MEDLINE REFERENCE	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)			
AUTHORS	3 (sites)			
TITLE	Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasekawa,C. and Shinagawa,H.			
JOURNAL MEDLINE REFERENCE	Complete nucleotide sequence of the prophage VT-1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak			
AUTHORS	Gene 258 (1-2), 127-139 (2000)			
TITLE	4 (sites)			
JOURNAL MEDLINE REFERENCE	Hayaishi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsuo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasekawa,C., Ogasawara,M., Yasunaga,T., Kuhara,S., Shibata,T., Hattori,M. and Shinagawa,H.			
AUTHORS	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12			
TITLE	DNA Res. 8 (1), 11-22 (2001)			
JOURNAL MEDLINE REFERENCE	5 (bases 1 to 266558)			
AUTHORS	Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayaishi,T.			
TITLE	Direct Submission			
JOURNAL MEDLINE REFERENCE	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kengen-info.osaka-u.ac.jp, URL:http://www.gen.info.osaka-u.ac.jp/, Fax:81-6-6879-2057)			
AUTHORS	genome project			
COMMENT	Location/Qualifiers			
FEATURES	1..266558			
SOURCE	/organism="Escherichia coli O157:H7"			
	/strain="O157:H7"			

gene /sub strain="RIMD 0509952"
/db xref="taxon:83334"
complement(244. .693)
/gene="ECs3650"
complement(244. .693)
/note="similar to B2790 ECOLI g1|1789154 percent identity 100 in 149 aa (Conserved in E.coli K-12)"
/codon start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="BAB37073.1"
/db xref="GI:13363122"
/translation="MARIGIFVGMVNGNSLVAEEARAIITAOGRKATVPEDELSDM
LPYODKVIIVYVSTTCGGDIPDSIVPIFOCIKDSIGCPMLRQGVLAIGSSVYNPCN
GGKQFDALDQESQRYGEMLDILASENPEPIESNPNVWQKTLIS"
complement(711. .1493)
/gene="ECs3651"
complement(711. .1493)
/note="similar to YOCB ECOLI g1|1789155 percent identity 99 in 260 aa (Conserved in E.coli K-12)"
/codon start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="BAB37074.1"
/db xref="GI:13363123"
/translation="MLEILYQDEWLVAVNKRSGMLVARSWLIDDEKVVYMQTVRQIG
QHVETARLDPTSGVILIMGISEAGLLAQEFOHQIQKRYHAIYRGVIMLEBAVDY
PLVREHLKIDAFKAREDKGPOPATVYTRGATVEMPVATQRYPTTRGYLVEPKQR
KHQIRLAIHRHP1IGDSSKGLDQNRSGAEHGLQRLMLHASQSLSTHPFTGSEPLT
IHAGIDPMQALQSFQWRGLPENERVEFSAPSGQGERSS"
complement(1493. .1822)
/gene="ECs3652"
complement(1493. .1822)
/note="similar to B2792 ECOLI g1|1789156 percent identity 98 in 109 aa (Conserved in E.coli K-12)"
/codon start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="BAB37075.1"
/db xref="GI:13363124"
/translation="NTHDRYRDLQALYLELRHQRWNRDEPQPNFTOPFPMT
MEFSLWDVWVILPRHDLNNQPLPAFAVATPEALDATHDRQRLILAELEKDA
LPAIDAS"
complement(1865. .2233)
/gene="cstrb"
complement(1865. .2233)
/gene="cstrb"
/product="cstrb RNA"
/note="inhibitor of CsrA"
complement(2444. .2989)
/gene="ECs3653"
complement(2444. .2989)
/note="similar to SYD ECOLI g1|1789157 percent identity 98 in 181 aa (Conserved in E.coli K-12)"
/codon start=1
/evidence=not_experimental
/transl_table=1
/product="Syd protein"
/protein_id="BAB37076.1"
/db xref="GI:13363125"
/translation="MDDITQALKDPTARYCODAMHEBKSMPLSEELYGVPSCIIIST
TEADVYMQPQEPTEGQNVNVERAFDPIVIOPTTHFTVTOGAPGMHAFQPDIKITLLO
TSDSDPRQVQENLIGHLVQKRLKPLTFILATLEELVIVISCNLSGVCKTGLTI
KXPHLADNLEFNLQKPLL"
complement(3905. .3905)
/gene="ECs3654"

KEYWORDS	SOURCE	ORGANISM
		Xanthomonas campestris pv. campestris str. ATCC 33913.
		Xanthomonas campestris pv. campestris str. ATCC 33913.
		Xanthomonas campestris pv. campestris str. ATCC 33913.

REFERENCE
1 (bases 1 to 10773)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., et al.

Quaggio, R.B., Monteiro-Vitorello, C.B., van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Carozzo, J., Chamego, F., Clapina, L.P., Giacarelly, R.M.B., Coutinho, L.L., Cursido-Santos, J.R., El-Doroty, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Maella, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melands, J., Merck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr. H.A., Rossi, A.L., Seta, J.A.D., Silva, C., de Souza, R.F., Spindola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezra, R.I.D., Trindade Santos, M., Tuffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Katsajima, J.P.

TITLE Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities

JOURNAL Nature 417 (6887), 459-463 (2002)

REFERENCE	2 (bases 1 to 10773)
AUTHORS	da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., et al.

Quaglini, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Carozzo, J., Chamego, F., Clapina, L.P., Ciccarelli, R.M.B., Coutinho, L., Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Fomichieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsunuma, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melanda, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A.J., Senna, J.A.D., Silva, C. de Souza, R.P., Spinola, L.A.P., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezera, R.I.D., Trindade Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.

Direct Submission
Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900,
Brazil

FEATURES	Location/Qualifiers
source	1. .1073

```
/strain="ATCC 33913"
/db_xref="ATCC:33913"
```

gene
79.1182
/gene="XCC1704"

/gene=accr07
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"

```

/label="CAOJECAT"
/product="hydrogenase subunit"
/protein_id="AM4098.1"

```

/ C l a i n s t a c i o n e m s d s f v l f r k g i s t a v l f a n k s v a e d f v e d l f r t a c c o n t a m e n t
 l l t v r k r g o n t a v a k a r l a o m a g i s t a v l k o r h a v t t o r f s v l p k r i a p d l
 a l d d m o v e h t w n r k l o r g a h g n r f v l t l r e v a d o a v i d a r l h a i l a a g c i p n n

RCSDADAIEQALSDPOSTLARITGEAAGIKQERALLRLRPOGLAHMWDQITQL

```

gene      complement(1439.1984)
          /gene="XCC1705"
cnc       complement(1439.1984)
          /gene="XCC1705"

```

```
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/locus_start=1
```

```

/product="conserved hypothetical protein"
/protein_id="AAM40399.1"
/def_name="AAM40399.1"

```

AERDELEAQSEFARLLRDSPLEAGTASYRRPTLPARLQRLKRGQESVQDELDLHG
ATAVQDALLROFLLEAHAHEHCVRILHGKGLQSDSGAPVILKNLVDRVLRQRDVLA
TAVQDQGGGGLTTRTANNH

```
/gene="surF"  
/note="XCC1706"
```

```
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
```

```

/-----
/product="survival protein"
/protein_id="AAM4100.1"

```

SLTLDVPIRTIRIDAQTCAVAGTPTDCYHALTGMLOYDPDIVSGINNSANLGDI
YSGTVAAMEGRFLGLPAVAVSLVTQNEAHHEFAARAANEIVARLKADPLPADTIL

```
gene 2854.3531
/gene="pcm"
```

```

2034: 1000
/ gene="pcm"
/ note="Identified by sequence similarity; putative; ORF

```

```

/coactm_escart=1
/transl_table=1
/product="L-isoaspartate protein carboxyl[methyl]transferase"

```

/protein_id=AA11100.1
/db_xref="GI:21112799"
/translation="MTPRLRLOPESVGIMTSQRYRDLVERLRRESGIQDEATLNAV

GÜGALLPALLUGLEVI I VERJEDURKUNARF RUDOMARKS KUNDGAT OG SINCE 1986
VTTAARAPALVDOLVAGGRUVAPVGASSSQSLVÖLTRADGEIAGLEVLAPEVEP
ILISGMID"

CDS
/gene="XCC1708"
350.4164
/gene="XCC1708"

```
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=1
```

```
/protein_id="AA041002.1"
/db_xref="GI:21112800"
/translation="MKRTGDIYDVATPSGSHPRADALITGISEVEGFTFPDPRVMIA
```

4143 4043
LQVVAESPWKAFMLLVI VGFETPIPLKFTWASGVGVPIEFIASMLVGRGRVYL
AGAIRLGGARAEALHRIEPIGWVAVLAVVAGMLVWKATNG"

CDS /note="XCC1709"
4143. .4943
/... "1-3"

located using Blastx/Glimmer/Genemark".

```

/codon_start=1
/translate=1
/product="lipoprotein"
/protein_id="AA041003.1"
/db_xref="GI:21112801"
/translation="MEGNERMSRAQMKLVKTAVMVLVVAAGLTACSSATVVRSPNSGS
GGASRPSAPRSPVPAPATVVRDITLVAISRNITAPLMAWNGLSNNTIYP
GGTLKLPSPAPSPGASAPVAVSGGTVVPPRPAIPAPVSGSGSMRPADGVVGFV
TGGTQGVDIAGASQAVRAADGVVAVGAGLVGELIITKHNDQMLSAVGNHRK
RLNBGSQVAGQIAEMGSRGAARMLHREINRGKPVPLIYPRK"
complement(5111..5485)
/gene="XCC1710"
/complement(5111..5485)
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/translate=1
/product="conserved hypothetical protein"
/protein_id="AA041004.1"
/db_xref="GI:21112802"
/translation="MPLNDSRDTYALRAADGRHAKVNSQILQSGRILMPDELVHM
PVSIGLOPRANDVAVLALNPAYITLGTGRQPPSTDVLAACLTGIGLEANTMAA
ARTYNVLAEGRRVVALMIVG"
complement(5502..5807)
/gene="XCC1711"
/complement(5502..5807)
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/translate=1
/product="conserved hypothetical protein"
/protein_id="AA041005.1"
/db_xref="GI:21112803"
/translation="MSIVTSAQNRPLRGANDLKLQTGKGVTPPAVLELDVLE
RHLIRVKAABERTRDALIAEIIHTGALVORIGHVALIYRPSKERQIVLPBG"
5867..6505
/gene="fcbJ"
/note="XCC1712"
5867..6505
/gene="fcbJ"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/translate=1
/product="cell division protein"
/protein_id="AA041006.1"
/db_xref="GI:21112804"

```

Alignment Scores:

```

Pred. No.: 0.000139 Length: 10773
Score: 234.50 Matches: 83
Percent Similarity: 40.73% Conservative: 40
Best Local Similarity: 27.48% Mismatches: 108
Query Match: 14.04% Indels: 71
DB: 1 Gaps: 11

```

US-10-018-706-2 (1-322) x AE012272 (1-10773)

```

QY 25 ILETHRTHRCYALLELEUALAGLYCYALASERYPROTHRYRAENSERTHRSERYGLY 44
DB 4224 CTGACGGCGCTGACGACGCGGACGTCGTCGCTCC-----AACTGGCGGACAGTGT 4277
QY 45 SERGLYSERHISARGTHRSEGLYSERGLYLEUALALE-----GLYSERGLN 61
DB 4278 GGTGGCGCTTCGGCGCCACTGTCGCGCCGTCCTCGGTGCGCGCTCCCGGTGCAACC 4337
QY 62 VALILETHRASPSERGINGLYVALPROANARGTYRCLINVALYSGINGLYASPTVAL 81
DB 4338 GTG-----ACCGTGGACGCGCGGCGACACGCTG 4364
QY 82 SERLYLEALAGLNARGTYRGLYLEUASNTTPARGTLNILEGLYHISILEASNALEU 101

```

```

DB 4365 TATGCATTTCCGCGCGCACCAATATACACGCGCGGACCTGCTCGTGAACGCGCTG 4424
QY 102 AANSESERLYRTHRIETRYTHRGINGINTPRLEUTHREUTHRESEGLYASPLELYS 121
DB 4425 TCTTCACCGAACGATCTACCCCGGACACGCTCAAGCTGATTCG----- 4472
QY 122 VALARGGLNARGSERILESESERGLYVALASNTHALAISTHPRSEPROVALA 141
DB 4472 ----- 4472
QY 142 VALGINSERSEARGPROPROVALINGLNHISPROALVALGINDLYSPROTHPROPRO 161
DB 4473 -----CCGCGCGGACGAAGCTGTGCAAGT 4499
QY 162 VALVALVALVALYLYSPROTHPROTHPROTHPROTHPROTHPROTHPROTHPROTH 181
DB 4500 GCACCGGTGGCGACGTGGCGGACGTCGTCGCGG-----CGCGCGGCGACCGATT 4553
QY 182 ALAPROPROVALINRGUALAPROPEALATHRGYSERSEGLYVALMETGLNPHARG 201
DB 4554 GCGGCGCGGCTC-----AGTACGCGG---TTCAGCTGCGGC 4586
QY 202 TYRPROVALGLYALATHRAENPROVALVALARGARGPHEGLYTHRALATHRVALAGLY 221
DB 4587 TGGCGG-----GCCACGCGCGTGTGTGTGGGACCTTC-----GTCAACGCGC 4628
QY 222 SERTHVALTHRESENGLYMETTRPHESEGLYARGASPTLYASPLEULEASNA 241
DB 4629 GAGACG---ACCAAGCAAGGCGTGACATTCGCGGTGCCAGCGGCAAGCGGTGCGGC 4665
QY 242 SEASNAHLAAGLYRTHVALILEGNAALASPHIASMECAASPTLYALASER----- 4759
DB 4686 GCGCGCGACGCGCTGTGTGTGTATTCGCGCGCGCGCTGTGCTACGCGGACCTATC 4745
QY 260 VALILEGNIETHRANGLYPHEVALSESELYRILEHISTILEYASAPALAGLNVAL 279
DB 4746 ATCATCAGACACACATCATGCTGTGTGTGCGCTGACGCGGCAACGCGGCGCTGCTC 4805
QY 280 LYSETHRGLYASPTLYRTHVALARGTHRGYGLNARGYLEALASMETLYASNGINPROSER 299
DB 4806 AACGAGGCGGACGCTGTGAAGCGCGGACCAATGCGCCAGATGGCGCGACGCGCGC 4865
QY 300 GLYALALEUAPHEGLUPHEARGILESERYASNGLYVALYRVALASPTLEUTHR 319
DB 4866 GCCCGGACATGCTGCACTTGAGATCCGCTACAGGCAAGCGGCTGACCGCTGCTG 4925
QY 320 VALLEU 321
DB 4926 TATTTG 4931

```

RESULT 16

AE004905/c 10822 bp DNA linear BCT 30-AUG-2000

LOCUS Pseudomonas aeruginosa PA01, section 466 of 529 of the complete

DEFINITION Genome.

ACCESSION AE004905 AE004091

VERSION AE004905.1 GI:9551195

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

20437337

PUBMED
2 10984043
1 (bases 1 to 109822)
Stover, C.K., Pham, X.-O.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,
Hickey, M.J., Brannan, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
Lagrou, M., Garber, R.L., Colby, L., Tolentino, E.,
Westbrook, M., Madan, S., Yuan, Y., Bick, J.L., Coulter, S.N.,
Folger, K.K., Kas, A., Lathig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reiter, J., Sater, M.H.,
Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
Location/Qualifiers
1. 10822
/organism="Pseudomonas aeruginosa"
/strain="PAO1"
/db_xref="taxon:287"
86..745
/gene="PA4918"
86..745
/gene="PA4918"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG08303.1"
/db_xref="GI:9951196"
/translation="MNRPSKIASPDVPAOKGPTPLCPDELPEVGDAIGGALNQLASR
AARIGSKDAHSPKAAWVPSASQDLARPLAPADLTWVSHCVGTPGPGFALDELPRP
IDYDFPKVGEEDLHPYGA CYHDLARSRSTGVTEYKONGSVHVLVGGALDYCVN
TAIQRRAGFEVLLYLPACRAIADPTQRACDNRDAGVILCDLELDALAGADGE
K"
748..1947
/gene="pncB1"
/note="pncB1"
748..1947
/gene="pncB1"
/codon_start=1
/transl_table=1
/product="nicotinate phosphoribosyltransferase"
/protein_id="AAG08304.1"
/db_xref="GI:9951197"
/translation="MAESVPAERIVQNLDTDFYKLTMMQAVLHNNYNAEWEFEFR
NAEDRLPYLAIRYQIERIAVEVTDQLAFLEKIPMKEDPTFLSLFFENLRYHT
GIEGGQAIIRLRGFWLHVIIFEVPLAIVSEVRNRYREVVLQVEQLYRKIDWLS
AQSSEELAEFOVADFGTRRRFSYRTOEVVHILKRGDPGRFVGTSMVHLAREYDLK
IGMAHEMLMAHQGLGRLVDSQQALDCWREYRGQIALTDCITMDAFLDFDLY
FALFDELRRHDSGDPILMAEKALHMYRLGIDPLSKTLVPSDGLMDPKALQLEFALG
KINVSFGIGTNLFCDIPGVHPMNIVLKMTACNGHPAKISDAPGKTCRDBNFVAYIR
HVNVPVA"
1976..2803
/gene="nabB"
/note="nabB"
1976..2803
/gene="nabB"
/codon_start=1
/transl_table=1
/product="NH3 dependent NAD synthetase"
/protein_id="AAG08305.1"
/db_xref="GI:9951198"
/translation="WQOIQRIQALQVQPPQSEADYVQACIARIATFOCCKDSGL
KTLVLGSGGVDSLTGALLQRAVEQLREQGDAYFTIARLPYQVQDADQASL
ATIRADEQVTNIGSPVKALAEQLAEGLPEPASDVFITAKIRIMVQAYIAGAR
GGLVIGTDHAAAEVAGFTKFGDAGCDLAPISGLAKQVRLARALGAPENLVKIFT
ADLEDLRPGHDEASHGVTYAEIDAPLHGQPLREBAARVIVDYHTKTOHKRELKPA"
2937..3860
/gene="PA4921"
2937..3860
/gene="PA4921"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG08306.1"
/db_xref="GI:9951199"
/translation="MRTRLPALLLGLVLLAGACGHTSPLLAFVROTHAFGDSYDNCE
SQRLEMLAKGIAGQALPGEVYQGRWNSGPTAVEVLAQOLADHDAVAGAKSG
ANRYGGMAYRHGTAGQGVDAVYATLDGKPVDAQALHIFVSANDFEHEDFAGQNP
LQIAGSSVANIIRAAVORGEAGARRFLVSSSTDLSVPAVAVGNRVERAQRVLQVNV
ASLPLOALAKRTRGELSMFPHLFSRILRNPNARYGVLEIDAPCPQPSVRPACA
NDQYFMEWHMPTRRVHQLAGEAAMARAR"
complement (3929..4375)
/note="PA4922"
/gene="azu"
complement (3929..4375)
/translation="MTRKAAVSLSLASAPLLAECSVDIQGDMQPNITATYDK
SCKQFTVNLSPGNLPKYNVGNHWSLADQGVVTDHAGSLDKDYLPDPSRVIA
HKLIGSGEKDVTEDVSKLKEBOYMEFCTPGHSLMKKTLTK"
4651..5238
/gene="PA4923"
4651..5238
/gene="PA4923"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG08308.1"
/db_xref="GI:9951201"
/translation="MTLRSVCEGCGSPGASPVYCEAAVALGRLAERGLTVYGGGA
VGLMGVADAAALAAAGGVIGIIPQSLQEAIEGKGLGLEVDGMAKARARMAELDA
FLAPGLGLTELELFEVMTWGOGLYAKPLGLELVGFYDPLITFLHLYDERFVAB
HRCMLORGASPEALDIALAMTPSVAPKVVDPG"
complement (5251..5946)
/gene="PA4924"
complement (5251..5946)
/gene="PA4924"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG08309.1"
/db_xref="GI:9951202"
/translation="MFRLLIIVLILAVSAGSKAPGASVRYKRYRGRVYRGDTL
YSIATHGNNYKOLARANGIRPPYAVKQVAFDERKSTYVASSSSSTRARKPP
PNSVATIRGNQWQMKGPVIRRESSDKLNKGIIRIAGTLQPVGASTLAGKVFANNR
GXENVYIRIGHGYSYSTYVHNSRLLVKEQWVGKQKIAAGSSDADRVQVLYEIRON
GRGLDLSLSPS"
complement (6133..6984)
/gene="PA4925"
complement (6133..6984)
/gene="PA4925"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG08310.1"
/db_xref="GI:9951203"
/translation="MDDTIVNVGTCKINALMTVSOYAFAGIKIILAAIAFVIGRW
LIGFAVGQKSLERKQVDPVTYRVGSFTVLNVLAVGVIGVKEGICTTSLAALIA
AVGLATGMAISGLANLAAGGFTIVARPFVGGFTIAGGTGTVKEGILETALIND
NVLTWGNKNTIRGDNINQPHNPFRRVELKRLSGADYQALATLAKORVAAIPNVIG
EPYDVYIEFLNVLGVLAVRPYCHNDHWQVYFDINKYIKDALGADFPAPMAQYVI
VOOSAKA"
complement (7235..8170)
/gene="PA4926"
complement (7235..8170)
/gene="PA4926"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG08311.1"
/db_xref="GI:9951204"
/translation="MSAPALRSVPLQGARVYVLAHDTYRYVAMFVSLSRQIAHLMREPC
AMACHERRALLIEPSPRRDGHDAFGNPLRIAFERPHALRSARLQVEVPAHAPAA

gene	complement (7237. .8320)
	/gene="STM3026"
CDS	complement (7237. .8313)

Alignment Scores:	
Pred. No.:	0.00041
Score:	232.50
Percent Similarity:	40.15%
Best Local Similarity:	24.41%
Query Match:	13.52%
DB:	1
Length:	22400
Matches:	76
Conservative:	79
Mismatch:	99
Indels:	87
Gaps:	9

US-10-018-706-2 (1-322) X AE008839 (1-22400)

[illegible]

```

Qy      311 AsnGlyValTyrValAspProLeuThrValLeu 321
          ::|||
Db      18220 CGGCGACGCGCATGCATCGCTGCTATCTG 18188

```

LOCUS	DEFINITION	Accession
AL627277/c	Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome; segment 13/20.	AL627277

SOURCE ORGANISM	Antigen
Salmonella enterica subsp. enterica serovar Typhi.	1
Salmonella enterica subsp. enterica serovar Typhi	2
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae	3
Salmonella.	4

REFERENCE AUTHORS

REFERENCE
AUTHORS

1 (Passes 1 to 230050)
Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D.,
Wain, J., Churcher, C., Mungall, K. L., Bentley, S. D., Holden, M. T. G.,
Sebahia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
Conerton, P., Cronin, A., Davis, P., Davies, R. M., Dowd, L., White, N.,
Barrat, J., Felwell, T., Hamlin, T. S., Haque, A., Hien, T. I., Holroyd, S.,
Ungels, K., Krogan, A., Larsen, T. S., Leather, S., Moile, S., O'Gaora, P.,
Perry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S. and Barrall, B. G.

JOURNAL	Nature 413 (6858), 848-852 (2001)
MEDLINE	21534947
PUBMED	11677608
REFERENCE	2 (bases 1 to 230050)

JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
COMMENT E-mail: parkhill@sanger.ac.uk

FEATURES

Details of *S. typhi* sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).
Location/Qualifiers

Source

```

Gene
309..1598
/gene="STY3109"
/note="sdac"
309..1598
/gene="STY3109"
/note="Pasta hit to Q46940 (409 aa), 40% identity in 416
aa overlap
Pasta hit to TDCC_ECOLI (443 aa), 50% identity in 442 aa
overlap
Orthologue of E. coli sdac (SDAC_ECOLI); Pasta hit. to
SDAC_ECOLI (429 aa), 95% identity in 429 aa overlap"
/codon_start=1
/transl_table=1
/product="putative serine transporter"
/protein_id="CAD02796.1"
/db_xref="GI:16504050"
/db_xref="SPTRMBL:082435"
/translation="MKTKQSTTASISRSBAMKRTDPMWGLXGTAIGAGVLPPLN
AGVGMDPLILMALPMPMTFAHRLGTRVLGSKNGEDITEVEESHFIIAGKAIT
LLFYFAPIPLLVAIVATNTVVESEFLTHOLAINPPRAISLILVGMNTIVFGEOM
IVKMSILVFEFAVAILMLALYLIPQNGALLETLSDAISGNGKIMLMLAIPLM
VESFNHSHTISFAVAKREEGEAKGKCKSLFAHIMHMLVMPPEVSCVSLTPA
DLAAKRNISILSYLANHRAPIIAMAAPILMAITATKSFLEHIGARGRGNMYS
SLRGKESIEINKNKTALFMVLTNTIVATNLSIGMETTGGPIIAITLPLMPY
AIOKVPARKSGHISNVFVIMGLNISAIIPSLSE"
```



```
gene      1657..3024
           /gene="STRY3110"
           /note="sdab"
CDS       1657..3024
           /gene="STRY3110"
           /note="Fasta hit to TDGC_ECOLI (454 aa), 73% identity in
           454 aa overlap
           Fasta hit to SDHL_ECOLI (454 aa), 77% identity in 455 aa
           overlap
           Orthologue of E. coli sdab (SDH_ECOLI); Fasta hit to
           SDH_ECOLI (455 aa), 94% identity in 455 aa overlap"
           /codon_start=1
           /transl_table=11
           /product="L-serine dehydratase 2 (L-serine deaminase 2)"
           /protein_id="CAD02797.1"
           /db_xref="GI:16504051"
           /db_xref="SPTREMBL:Q82434"
           /translation="MISVFDFIKIGIPSSSHTVGEMKAKGKFTDILARHIIIDVTR
           VVDVYSGSLTGKGHITDIALIMGLAGNLPDVIDALPSFIQDVNTHGRLLANGQ
           HEVEFVDKCMNFHADNLSLHENGRIITALAGKRVLSQTYSGIGGFVDEBFGILT
           NSEPVAVPYKYTAADLORHCETGLSLGLMMQNEIALHSKEALOECHFWKWEVMS
           GIERGITSGVLPGLKRVPRRAALRMVLSQDKTTDPAVVDWIMFALVNEENA
           AGRVVTAFTNGACGIVPAVLAYDKFIREVNANSLARVMLVASAIGSLYKMAISG
           AEVCGGEVGVACSMAGIABLLGSPAVOCIAEIMAHENIGLTCDDPAQGVQVPC
           IERNALASVAVAAARMALERTSEPRVCLDKVETWYETGKDMAKYRETSRGLAMK
           IYVCD"
gene      3135..3950
           /gene="STRY3111"
           /note="exo"
CDS       3135..3950
           /gene="STRY3111"
           /note="Orthologue of E. coli exo (EX9_ECOLI); Fasta hit to
           EX9_ECOLI (281 aa), 88% identity in 264 aa overlap"
           /codon_start=1
           /transl_table=11
           /product="5.3' exonuclease"
           /protein_id="CAD02798.1"
           /db_xref="GI:16504052"
           /db_xref="SPTREMBL:Q82433"
           /translation="MLASPRRECTPFRPREYRENTVAHLIIVANLIRIRIHVQSP
           CVTECOALDQLIHISQTHAVAVFDDASSGSGRQRLDTPKGRPEPPDLDLNEP
           ALTAPEFGRCVCMASDGNEDLAATLAKTBAGQATIVSTDKYCSLSEGLTI
           RYDFQKRLDAPLEKEFGLVPRQLPYWGLAGISSKSPGAVGIGKXSTQLVQPO
           NREGIYAHDEVEPEKMKLETHKEMAFICRDIAIRQTDLIDGNLQQLRLVR"
misc_feature 3201..3947
           /gene="STRY3111"
           /note="Pfam match to entry PF01367 5.3' exonuclease, 5'-3'
           exonuclease, score 120.40, E-value 3.3e-32"
           complement(4042..5190)
           /gene="STRY3112"
           /note="fucO"
           complement(4042..5190)
           /gene="STRY3112"
           /note="Fasta hit to EUTG_ECOLI (395 aa), 40% identity in
           372 aa overlap
           Fasta hit to Adh2_ECOLI (383 aa), 44% identity in 371 aa
           overlap
           Orthologue of E. coli fucO (FUCO_ECOLI); Fasta hit to
           FUCO_ECOLI (383 aa), 91% identity in 382 aa overlap"
           /codon_start=1
           /transl_table=11
           /product="1,2-propanediol oxidoreductase (lactaldehyde
           reductase)"
           /protein_id="CAD02799.1"
           /db_xref="GI:16504053"
           /db_xref="SPTREMBL:Q82432"
           /translation="MANRMILNETAWMGKGAVALDTEVTRGYHKAIIYTDKTLVOC
           CKVGVTSRMDAAGLWEIYAGVIPNPITISVOGLKVFQSGADYIIAIGGSGPQRT
           CKVAGIISNNPEADVRLSEGLSPTKRPSPVIMAIPTAGAAEVITINIVITTBESR
           KFCVCDPHDIPQAALFIDADWDKMPALPKATGVDALTHAIEGVTAAVALDIALHI
           KALEIIGALRGAAVAGEKEGEMALQOQVYAKMGFSNVGLGVGMHAPGAFNTPH
           GVANAIIIPVMEFNRAGSTNEKFRDIAIRANGVKEGSLSEARNAANAEEAFTLNRDVG
           IPILRDVGVRKEDIIPALAAAFDVCCTGCPREASLADIHIVELVHTAW"
```

```
misc_feature      complement(4069..5169)
                   /gene="STRY3112"
                   /note="Pfam match to entry PF00465 Fe-ADH, Iron-containing
                   alcohol dehydrogenases, score 655.70, E-value 2.4e-193"
                   complement(4354..4416)
                   /gene="STRY3112"
                   /note="P800060 Iron-containing alcohol dehydrogenases
                   signature 2"
                   complement(4585..4671)
                   /gene="STRY3112"
                   /note="P800913 Iron-containing alcohol dehydrogenases
                   signature 1"
                   complement(5207..5854)
                   /gene="STRY3113"
                   /note="fucA"
                   complement(5207..5854)
                   /gene="STRY3113"
                   /note="Fasta hit to YGBL_ECOLI (212 aa), 34% identity in
                   179 aa overlap
                   Orthologue of E. coli fucA (FUC_A_ECOLI); Fasta hit to
                   FUC_A_ECOLI (215 aa), 92% identity in 215 aa overlap"
                   /codon_start=1
                   /transl_table=11
                   /product="fuculose-1-phosphate aldolase"
                   /protein_id="CAD02800.1"
                   /db_xref="GI:16504054"
                   /db_xref="SPTREMBL:Q82431"
                   /translation="MERNRRLARQIIDTCLMTRRLGNOGTAGNSVRYOGMLITPFG
                   IYEXKLTBSHIVFIDADQGEQGLPSSSEMRHMAVYQTRPPANAVNHNHVAHCTAVS
                   IINRPIPAIHVMIAGAGNSIPCAVATPRTRELSDBHVALKMRKATLLOHGLIAC
                   ENRIDKALMAHEVEVIAQLVSTALVDPVPLVDEALVIALEKRYGLRIEE"
                   complement(5219..5845)
                   /gene="STRY3113"
                   /note="Pfam match to entry PF00596 Aldolase II, Class II
                   Aldolase and Adducin N-terminal domain., score 337.10,
                   E-value 2e-97"
                   6410..7726
                   /gene="STRY3115"
                   /note="fucP"
                   6410..7726
                   /gene="STRY3115"
                   /note="Orthologue of E. coli fucP (FUCP_ECOLI); Fasta hit
                   to FUCP_ECOLI (438 aa), 92% identity in 438 aa overlap"
                   /codon_start=1
                   /transl_table=11
                   /product="L-fucose permease"
                   /protein_id="CAD02801.1"
                   /db_xref="GI:16504055"
                   /db_xref="SPTREMBL:Q82430"
                   /translation="MGNNTTIQTSFRAVDAEBSKRRYIIPFALCSLFFLVANNL
                   NDILPQFOAFLITNFQAGLIQSAFVGVFVPIPIAGILMKLSKAGITGLFLYA
                   FGALFWPAAEINNTYTLFLIGLFIAGAGCETANPNPVALKSGSGHFRNLNLOT
                   FNSFGAIIIVAFEGOSLILSNVPHOSQALDKMTDPDLSAKSLVSVQTPVMTIVAI
                   VIVVALLIMLTKERPALOSDSDHSPAKQSSTLSLSRLIRLRHMWAVLAQFCVGAQTA
                   CMSYLIRYVIAIEBIPGMTPGFPAANYLTGTWVCEFTGFTGTLISRPAPHVLAAYAL
                   AMILCLISAFSGGSHIGLALTLCSAFMSIOYPTIFGLKTNLQGDTRKYSSFTVMTII
                   GGGIVTPVWGFSVDAAGKIPTELVPALCFVAFIFARRRQATN"
                   7758..9488
                   /gene="STRY3116"
                   /note="fucI"
                   7758..9488

gene      0.00535
           /length: 230050
           /score: 232.50
           /percent_similarity: 76
           /best_local_similarity: 40.19%
           /query_match: 24.44%
           /mismatches: 99
           /indels: 87
           /gaps: 9
           DB: 1

US-10-018-706-2 (1-322) x AL627277 (1-230050)
Qy      18 LeuGIyLeuIlePheGIyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysPro 37
```



```

Db      84819 CTGGGATTGCGGTGCTGATATGACAGCGGCTGCTGCGGGGTGT----- 84772
Qy      38 ThrThraSerThrSerGlySerHisArgThrSerGlyGlyLeuAla 57
Db      84771 -----TCCAGTAATTCGGCTCAGGAACGAT-----TCCGGCTCC----- 84736
Qy      58 IleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValIleGln 77
Db      84735 -----GTCATACCGCTTAACGC 84718
Qy      78 GlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTyrArgGluIleGlyHis 97
Db      84717 GCGGATACGTTGATGATATGCGCGCTACGGGAACCGCTGAAGAAGACTGCGCCGA 84658
Qy      98 IleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnThrLeuThrLeuTyrPse 117
Db      84657 CTGAATGGCAATTTCTCCGCTTACTATTAGAGTGGGCGAGGAATTAAGT---CGC 84601
Qy      118 GlyAspLeuLys-----ValArgGluArgSerIleSerSerGlyValAsnThr 133
Db      84600 GCGAGCGCAAGAGTAGCACTTCAACGCAAAACGCAATTAACAGCAACAAACT 84541
Qy      134 AlaHisThr---ProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152
Db      84540 GCTGCGGTCAGACCGTATCTCGGTGCGCAAAATCTCTGCGCGCGCGTGGGCGACGCT 84481
Qy      153 ProAlaValGlnLysProThrProProValValValLysLysProThrProThrPro 172
Db      84480 TGTGGGTGGTGGCTCTAACGGTAAAGTATTTTA----- 84445
Qy      173 ProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThr 192
Db      84444 -----CTTATTCACC 84433
Qy      193 GlySerSerGlyValMetGlnPheArgTyrProValGlyValThrAsnProValAlaArg 212
Db      84432 GCGGAGGCGCGA----- 84421
Qy      213 ArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTyrPheSer 232
Db      84420 -----AATTAAGGAGTATTTGCG 84400
Qy      233 GlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHis 252
Db      84399 GCGCGCGCGGAGACCGCTGTTATGCGGCGGAGCGGTAAGTGTCTATGTCGGGAC 84340
Qy      253 AsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheAlaSerSer 270
Db      84339 CAGCTCGCGGCTTAACGCAATCATCATGATTAAAGCAATGAAGTTTACATTACGCGCT 84280
Qy      271 TyrIleHisIleIleAspAlaGlnValIleThrGlyAspThrValAlaArgThrGlyGlnArg 290
Db      84279 TACGCTCATATGACACCATGCTGTTAATATATGCAAAAGCGTTAAAGCGGCGCAAAA 84220
Qy      291 IleAlaSerMetLysAsnGlnProSerGlyValAlaLeuPheGluPheArgIleSerArg 310
Db      84219 ATCGCTACGATGGGACGACGAGATGCCGATCCGTAAGACTTACATTCCGCTAC 84160
Qy      311 AsnGlyValTyrValAspProLeuThrValIleu 321
Db      84159 CCGGCGACGCGCATCATCGCTGCGCTATCTG 84127

```

RESULT 19
 AF260132
 LOCUS AF260132 1522 bp DNA linear BCT 16-MAY-2000
 DEFINITION Pseudomonas putida putative protein-L-isoaspartate
 O-methyltransferase (pcm) gene, partial cds; and NlpD (nlpD) gene,
 complete cds.
 ACCESSION AF260132
 VERSION AF260132.1 GI:7839529
 KEYWORDS Pseudomonas putida.
 SOURCE

ORGANISM Pseudomonas putida
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 REFERENCE 1 (bases 1 to 1522)
 AUTHORS Ojangu,E.-L., Tover,A. and Kivisaar,M.
 TITLE Sequence of Pseudomonas putida nlpD gene (complete sequence) and
 ORF similar to E. coli pcm gene (partial sequence)
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1522)
 AUTHORS Ojangu,E.-L., Tover,A. and Kivisaar,M.
 TITLE Direct Submission
 JOURNAL Submitted (24-APR-2000) Genetics, Institute of Molecular and Cell
 Biology at Tartu University, Riia 23, Tartu 51010, Estonia
 FEATURES
 source
 1..1522
 /organism="Pseudomonas putida"
 /strain="KT2440"
 /db_xref="taxon:303"
 <1..326
 /gene="pcm"
 <1..326
 /gene="pcm"
 /note="pcm"
 /note="methyltransferase"
 /codon_start=3
 /transl_table=11
 /product="putative protein-L-isoaspartate
 O-methyltransferase"
 /protein_id="AF70310.1"
 /db_xref="GI:7839530"
 /translation="RRSIDLHRAKERLVELNRYVFRWGDCBEGMPALAPYNGITV
 TAVAPVPOALLDQLAPGGRMVIPVPAEAOQMLIVREHGFSSRRVIGAVFVPL
 NGPLA"
 631..1440
 /gene="nlpD"
 631..1440
 /gene="nlpD"
 /note="lipoprotein"
 /codon_start=1
 /transl_table=11
 /product="NlpD"
 /protein_id="AF70311.1"
 /db_xref="GI:7839531"
 /translation="MGSLVGCSTSSNSARVVDNRNNTYKRPATVSGOYIVKPGDTL
 FSLAPRYGMDYKELARNGIOAPYTRIPGQAIRPSSGSSRTTVVSPSSSTTVTR
 RPYGSTRATPASTSKRATSPAPAPVATVPAPBAVAGVGTTPANGVILGRASNS
 LNKGIDVAGLDGPVPASDGAIVAGSGRGELITIKHSPTYSAVGHNRLLVR
 EGQVVRAGSILKMYVGTDRVKLHFIRROETRRSTVPATSLTVPSLFL"
 BASE COUNT 274 a 438 c 486 g 324 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,42e-05 Length: 1522
 Score: 231.50 Matches: 82
 Percent Similarity: 41.96% Conservative: 38
 Best Local Similarity: 28.67% Mismatches: 101
 Query Match: 13.86% Indels: 65
 DB: 1 Gaps: 12
 US-10-018-706-2 (1-322) x AF260132 (1-1522)
 Qy 20 leuIlePheGlyValIleThrThrCysIleLeuAlaGlyCyAlaSerLysProThrTyr 39
 Db 613 CTGGTATTCATGATGGCCATGGCTGCTTCTGCTGGGTTGCTCAAGC----- 660
 Qy 40 AsnSerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGly 59
 Db 661 -----ACCAGCAGCAAC-----AGC 675
 Qy 60 SerGlnValIleThrAspSerGlnGlyValProAsnArg----- 72
 Db 676 GCGCGGTGTCACCGCAACACACCGCCAGCCGCGGTGATCTTGGGCAA 735
 Qy 73 TyrGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTyr 92

[illegible]

REFERENCE 3 (bases 1 to 2609)
AUTHORS Seshadri, R. and Samuel, J. E.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2000) Medical Microbiology and Immunology, Texas A and M Health Science Center, 407 Reynolds Medical Bldg., College Station, TX 77843-1114, USA
REMARK Sequence update by submitter
COMMENT On Jun 1, 2000 this sequence version replaced gi:7340284.
FEATURES
source location/Qualifiers
1..2609
/organism="Coxiella burnetii"
/db_xref="taxon:777"
/note="nine mile phase I"
1..777
/note="function unknown"
/codon_start=1
/transl_table=1
/product="SufE"
/protein_id="AA073516.1"
/db_xref="GI:8141682"
/translation="MKRTAFPKRLILSNDGVYAKGLATIAKTLADLGEVDVYAPPR
NRSASVGLTVAALAEKRGFLGALPAAVSGELGFRYTEAKVYQILQIRDPDLE
NLGDVYSGTVAALAEKRGFLGALPAAVSGELGFRYTEAKVYQILQIRDPDLE
PSTLILNVDPDLEPEELKGFVTRLGTRHAEPIRQIDRGNFTIYVGAAGFQDSE
PGDFFAMNHCVSITPLRVDLTHYFAFDQASVWKLKM"
779..1471
/codon_start=1
/transl_table=1
/product="lipoprotein precursor NlpD"
/protein_id="AA073517.1"
/db_xref="GI:8141683"
/translation="MKRCIVVILIAFLICGCVESKNAPVNVGMLOPKRALGSRVVK
GDITSLAMAFGLDPRALAAANLSPYREIAQGLQMTIIPRAHVSGRPAAPRWK
SPPEQVAMRWMPAKGRLIGYGVGMAAGNGINIAHGIVAAVSAASGAGVSGAG
RGYNLILVKKHTNYLSAVAFNKKVLVEGSRVAVAGOKIAMGRTNSGRVWLHFEIR
NGQVNPFLRYS"
1551..2609
/note="global regulator of gene expression"
/codon_start=1
/transl_table=1
/product="stress response sigma factor Rpos"
/protein_id="AA061172.1"
/db_xref="GI:7340285"
/translation="MKTKTKTKTKIKKAAKKIKKPSKKRIKTKAKSRPKPKIKASDHG
LIPKTKETKEKEDALANAKKTKRRETRSDPQIYLRGFGOPFLNAKEELK
ABRPKGDPRKAKROMITANLAIYKIRAHVNRGLPFLDIEBGNIGLITVEKDDP
RGPFSTYATMTIRGTEIRALMOSPTPLPIHYIKELNYLPTAKKLTQVDEHAT
EDVAILDKPVOEIRIMDLAPSKSTIDVPSDQGSQVITADNNIIDPARLQNN
DLQDHIERWALQDDBRRREVITLRFGLHENEKGTLEAVGKAVGLTRRIVRQIQIDALC
QLRHIERWEGTVEVED"
ORIGIN 730 a 568 c 676 g 635 t
BASE COUNT
ORIGIN 730 a 568 c 676 g 635 t
Alignment Scores:
Pred. No.: 7.56e-05 Length: 2609
Score: 227.50 Matches: 82
Percent Similarity: 38.61% Conservative: 35
Best Local Similarity: 27.06% Mismatches: 101
Query Match: 13.62% Indels: 86
DB: 1 Gaps: 9
US-10-018-706-2 (1-322) x AF244357 (1-2609)
QY 24 ValIeThrTrhCysIleLeuAlaGlyCysAlaSerIys-ProThrTyPAsSerThrSe 43
Db 800 ATcATTATCGCTTCTCTGTCGGCGGTGTGTGAAGAAACAAAAATTTGCGCGCTTGGC 859
QY 43 rGlySerIysSerHisAspGlnSerGlySerGlyValIyLeuAlaIleGlySerGlnValI 63
Db 860 AACGCGTGGCTTCAACGACAGGCTTA-GGTTCT----- 895
QY 63 eThrAspSerGlnGlyValProAsnAspTyGlnValIyGlnGlnGlyAspThrValSerIy 83

[illegible]

gene homologous to E. coli K-12 MG1655"
complement(2289. .2362)
/gene="glyu"
/note="24202"
complement(2289. .2362)
/gene="glyu"
/product="rRNA-Gly"
/note="anticodon: "CCC"
/function="rRNA: tRNA"
complement(2487. .3156)
/gene="24203"
complement(2487. .3156)
/gene="24203"
complement(2487. .3156)
/function="putative structure; Not classified"
/note="Residues 1 to 192 of 223 are 96.95 pct identical to
residues 22 to 213 of 259 from Escherichia coli K-12
Strain MG1655: B2865"
/codon_start=1
/transl_table=1
/product="putative lipoprotein"
/protein_id="AAG57994.1"
/db_xref="GI:12517383"
/translation="MMLLSVGLLLAGCSGKSSPDYTGYSGVTVKRGDTLYRISRTT
GSVPEVGLANIGSPPTYLEVGOKLGGAKSSSTSTKSTAKSTTKTASVYSSAVDK
GSMPEVGGCWMPLPTTGKIMPEYSTADGNGKIDISARGTPIAAGAKGVVGNGLD
PCNGNLMIKHSEDIYITVAHNDTMLVNNGOSXEGMAKRRHGHGCGICSPAPDLSL
PCNGN"
3652. .5910
/gene="24205"
3652. .5910
/gene="24205"
/function="orf; Unknown function"
/note="Residues 1 to 752 of 752 are 99.46 pct identical to
residues 1 to 752 of 752 from Escherichia coli K-12 Strain
MG1655: B2865"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG57989.1"
/db_xref="GI:12517384"
/translation="MEVDLAKVTGRARYTDDYVMAAGCYAKYRSPRIAGYVNSIND
EQRSLREGVALITMEVDPLDIPFATAGHMTDESKRTADRLITRVRHNGAVAI
VQADDLITKKAOLASISEBELPVITIPALBEDAPVHNGGLAKQSTNSVGNVO
OTIDADYVQYOTPVQIRCHMESVYSLAMEDEDITTVSSQIPIHTRVAVGGA
DLIPMSCAVITFVGGGFGKQDVLSEPMALFSGIKGIPKVSISRECFATAT
RAFTIDCGVAVNDGDTLKYSLDIVLSNTHIASAGAKVAAYLPRCAVA"
SSKTCYTNIPSAAGKMGYGAQPVFAVSSMDDAATYALGIDPEIRLNARSEDNAP
LTKRTYSLSGPELCEKRGKRIPEWERRAECONQGNLRGVGACFSYTSNTPVAV
ELAGALLNNDQGITINVOSGATEIGQADYFOSQVAVTGVPSVDRAVISTQTDVAT
PDPGAFARQGSYVAAIPALRSALILKRIKIIAAHVAWLHGSAMLTLLIKGIIIVEER
BELMSIKLMDLADFAHPERGOQLSESIKTTNPPAFGTFVDLITVDLILCVTIN
RLNVLDSCHILNPLAEQGVHGGMGVIGMLFPEMTIDKSGVRRPNLLDVKMPTF
MDPLPLEAEFAFEINEPQSGAYHKSLGSEPPITPVAAIRNAVKMATGAVALITLPEK
RLYEERHLAGLI"
5921. .6799
/gene="ygeT"
/note="24206"
5921. .6799
/gene="ygeT"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 98.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655: B2867"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57986.1"
/db_xref="GI:12517385"
/translation="MEDFASYRAATVLADAINLNDNPQKILAGCTVLIQLHHND
KRIHVDIHNILKRGITLLAEDGSLRIGSATTFTXIDESTIQRHLPLCAASSTIKG
PQIRNAVYDGNINCATSDASATPLLIDALCLHISPRGVFVINFHPGPKVS
EADILVAHPPEQPEHVGSAHFYKAMDAEDISTICCAACRINDNSRELLRLARG
VAAPITVRCOHAQTAQNAPLNDIQTLELSSVDLVDRSSWRASKEFRLILQTMV

[illegible]

[illegible]

COMMENT Christian.Bouchet@coulouise.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES
source
1. .207050
/organism="Ralstonia solanacearum"
/strain="GM1000"
/db_xref="taxon:305"
complement (332. .1561)
/gene="RSC1148"
/note="RS04604"
complement (332. .1561)
/gene="RSC1148"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CADI4850.1"
/db_xref="GI:17428162"
/translation="MNELPLPYAARGVQPGLAGRPAQLGGSPDPMPRTMOKPL
LIRQALSPEDMRALHPPLSPDALIKLARREDSRLIAQRGRWTFNHPNERPLPS
RKAMNTLVQGVNVLVSPAVDALMQRPRLPDALDDVMISFATDGGVGVPFSYVS
FLICQHGRRMRIRISSQNDLTLPDLPLKILANFPPEEFILPECDMYLPPHYADGD
AEGCGHVSVPGRSPSYRRELAGHPLGFSQLEONPDPEGRYTPDEOKYVRPEGL
PMALAKOKNLAKMPLWPELVGEFPGAYLTPKQHVETVTPRLSLAFTRARREGIVL
LDARTALYDAQRWINGDTPFEPSDTLANLSALSDRGASAEAVDAADLPDMDITL
HAWEEGMLQLAGGAH"
1573. .2097
/gene="slyD"
/note="RSC1149; RS04602"
1573. .2097
/gene="slyD"
/EC_number="5.2.1.8"
/function="cell processes; chaperoning"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="PROBABE FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS
ISOMERASE (ROTAMASE) PROTEIN"
/protein_id="CADI4851.1"
/db_xref="GI:17428163"
/translation="MKIAKNTVVSVAKLSDAQNVLEESDEPMVYLHGVDGTFPKK
EALIDHEOGFEFTKLOLEPDDAGEYVQELIKLEPRRPPEPLEVGMQFEGVPGDGD
EDALITYVTVDVAEKVVLDGNHPLAGNALFDRLVYDVREATAEIEIHGHVHNGGLE
VADDEEPRTIIL"
complement (2152. .3216)
/gene="RSC1150"
/note="RS04601"
complement (2152. .3216)
/gene="RSC1150"
/function="miscellaneous; unknown"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="PROBABE TRANSMEMBRANE PROTEIN"
/protein_id="CADI4852.1"
/db_xref="GI:17428164"
/translation="MARFGAIFRAMKGIQAALMLATAPAPAKGHATAPAPACF
TFAVGVNPPRPEDVAPARALDLIDIEHPAFVVIHGNLKGSDSCITDNLIEEHDLL

DSTRTPYIYRGHDSMDCGRPTAGFDPPIERLRLADLPYDDNALGORTMYWROS
 DOAKRSPYRENARNGCVLFTLNNYNNYPTACGRNGEYVDRLEAKRQWYARF
 AIAQRRLPCQVITMOMADPLFEDMERRRARNLIDGLRRPHDGYALAKROLALAK
 EIDGAVLLINGASQSFIDRPLKDSGSPDHVAKRATFSGPTIDOMLEVSVTFGAP
 LFRHARRASDIDASTPOQ
 complement (3216...5864)
 /gene="muts"
 /note="RSC1151; RS04600"
 /complement (3216...5864)
 /gene="muts"
 /function="macromolecule metabolism; macromolecule
 synthesis, modification; dna - replication, repair,
 restr./modif."
 /note="Product confidence : probable
 Gene name confidence : probable
 predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not experimental
 /transl_table=1
 /product="PROBABLE DNA MISMATCH REPAIR PROTEIN"
 /protein_id="CAD14851.1"
 /db_xref="GI:17428165"
 /translation="MAHSAAPALEIDPKYGPDPKHTPMOQYLKLSAHPHLLVYEM
 GDFEYLPEDAEKARLLDITLRSQTTNGQIRMAQVPPHVAQVYALAKVGEVVA
 IGRQIDPATTKGVERKVVVVTGTLTDAALSDVNNHLLAIQIPGRGAAPV
 GLAMINVGGELRIMESABOLDELEIRIPAVYALDADATNAVAPVARRRLDMH
 DVENAGRIARROLGVASIVARGCETITPAALNLYAALNVAQGLHPIGICLVENH
 ESERIGDTPARVLELTERRCOESPPLRLLPTCCANMSGRILRHHHMLHLPDRAV
 PORAQALIVILLADKMSIRGTLRSTIDVERTRIALISRPDISSTIDPTLIR
 VRQDLPSSDDQLGDAHAALIPEDHAALQRYVAPAMPVDCQVATRGYARD
 ELRDISNCGVFLDLKEREKRTQANLREKRYVCFTEVINGCAKPRVYRRD
 QTLKASERTITPELAFEDKLSQADREKYLITSDQLPHLPVLRLLAALA
 QADVATLAEARALMSRPALVTPAGLELRAPHVVEQVEQFVADCVLQETKIL
 LITGPNMGKSTFMRQYALVLAAGAFVPAADATIGPIDRIFRIGADNDLGGK
 STFWEMTEAAAILHRAIPNSLVIMDEIGRTSFDGLAMALAHLLSHNSHTLE
 ATHVFELTQLEPQEAQANVHLSVEHGDGVFLHVAQEGPASQSYGLQVQLAGVQ
 VPIVARRRRLAMELOHSDATGATPOLDFALPSPSDDAEAAPAPASALAEALQD
 IDPDMTPRDALDLYRLKALSDASA"
 complement (5960...6853)
 /gene="RSC1152"
 /note="RS04599"
 /complement (5960...6853)
 /gene="RSC1152"
 /function="miscellaneous; not classified regulator"
 /note="Product confidence : probable
 Gene name confidence : hypothetical
 predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not experimental
 /transl_table=1
 /product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"
 /protein_id="CAD14851.1"
 /db_xref="GI:17428165"
 /translation="MKTLEELAPRYVVDGVTAAAEQGVVSGVSRALRLEK
 LETLLSRTIRRLADBGTAFLNARKIILAVBAEELILNRQOPRGRIRVNAAP
 FMHVIAPLAVDFRQVPOILELNTSRITIDLESDVDAIRIGALDSTLRARIG
 TRRLVILASPAVLAHGRKPTVDALRHALLIGFVOPSLNMPLAGPHGDLRTPTL
 KASSGETIRQLAGAGICLADPMTADRAGDLYOVLPEATVETIQPIHAYVYRT
 QASRITCFDLVLEARLAQGE"
 complement (5966...8168)
 /gene="RSC1153"
 /note="RS04598"
 /complement (5966...8168)
 /gene="RSC1153"
 /function="cell processes; transport of small molecules"
 /note="Product confidence : putative
 Gene name confidence : hypothetical
 predicted by Codon usage
 predicted by Homology

predicted by Framed"
 /codon_start=1
 /evidence=not experimental
 /transl_table=1
 /product="PUTATIVE TRANSMEMBRANE EFFLUX TRANSMEMBRANE
 PROTEIN"
 /protein_id="CAD14851.1"
 /db_xref="GI:17428167"
 /translation="MPTALALTLSPALCTTPEYVGLIPYAADIGTILPSAGNLY
 SLVAGVAGAPVLTATGRLPKRALALMALFTIGNLAMAPGTAIPVLAQWNG
 LAHVGFPSIGTATSLVPERKASAIATFTGLTVALVGVPLGTIGHFKMRETF
 LAVAGGVIAFVGSLSLVFPHDIAHAPPASARLOLYOQRLNVAKTIAIGVGSFT
 PFTFLVETLQDVAGFSAGVWMLVGVSAQVAGNMGHRLADRKGPITALQVTFALL
 AAVLFAETFTARYPMTAVATVLLMGAVAFNVGLQAVVQKMRFPAPQADVASGLN
 IAAENPGLAGAAWAGWIVTHGLMHPWIGALVIVLSGLTAMSGALDRGGIGPRL
 DGPAPVGH"
 8165...9016
 /gene="RSC1154"
 /note="RS04597"
 8165...9016
 /gene="RSC1154"
 /EC_number="1.-.-.-"

Alignment Scores:
 Pred. No.: 0.0186 Length: 207050
 Score: 222.50 Matches: 81
 Percent Similarity: 39.38% Conservative: 45
 Best local Similarity: 25.31% Mismatches: 135
 Query Match: 13.32% Indels: 59
 DB: 1 Gaps: 9

US-10-018-706-2 (1-322) x AL646063 (1-207050)

Qy	8	AsnSerGlnAenglnlySprolielysArgLeuGly	-----LeuIlePheGlyValIleThr	26
Db	70656	AGCTCCGTAACCTCCAGACCCCGCGCGCGCGCTCGTGCAGATGATCTCG		70715
Qy	27	ThrcysIleLeuAlaGlyCyAlaSerLys	-----ProthrTyraSenSerThr	42
Db	70716	GCCGCACTGCGGGGCTTCGCGCACTCCGCGCAACGAGCGCGCTCGTGCAGCAG		70775
Qy	43	SerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnVal	62	
Db	70776	TCCGCTCCGCGCACT	-----GCCCGCGCC	70799
Qy	63	IleThrAspSerGlnGlyValProAsnArgTyGlnValIlysgInGlyAspThrValSer	82	
Db	70800	GCGCGCTGGAACCGCTCCGCGCGCTACTATCGGCTCAAGCGTGCATACCGCTTAC		70859
Qy	83	LysIleAlaGlnArgTyGlyLeuAsnTrpArgGluIleGlyHisIleAsnAsnLeuAsn	102	
Db	70860	AGCATTCGCTACGCAACCGGCAAGCGCGCGCATCTCGGACGTGAACCAACATCCCC		70919
Qy	103	SerSerTyThrIleTyThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeuVal	122	
Db	70920	AACCGCAACGATCGAAGTCGACCAACTGATCCGCGT	-----GTG	70961
Qy	123	ArgGlnArgSerIleSerSerGlyValAsnThrLahIsthrProSerProValAlaVal	142	
Db	70962	CCGCCGAATGGACGCGCAGCGCGCAGCGCGCGCTCGCGACGCGG	-----	71006
Qy	143	GlnSerSerArgProProValGlnGlnHisProAlaValGlnIlySprothrProProVal	162	
Db	71007	-----GTCCGCGCGCC	-----	71018
Qy	163	ValValValIlySlySprothrProthrProProValValGlnGlnProAlaProValAla	182	
Db	71019	-----AATACACACGACGACGACGATCGATCGCGCGCGCGACGCGG		71063
Qy	183	ProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTy	202	
Db	71064	CCCGTGTCTGTCTCGCGCGCTTCTCGCGGCGCAACGACGCGCATCGCATCGCTTGG		71123

Oy		203	ProVal1g1y1a1eThraenP-roVal1Va1ArgAxpPhcG1YThr1aThrVal1aIag1yser	222
Db		71124	CCC-----GCCATGGCCACAGGTGAATGCAGCGCTTC-----GAC	71156
Oy		223	ThrVal1ThrSerAnGIyMetTriPhSeSerGIYAArgApG1YAsePleu1leaMa1aser	242
Db		71157	GACAAGGCCAAAGAAGGCATCATGATCGCGGCAAGCGGGCGACCGGGTCAACGGCTGCC	71216
Oy		243	AsnAlaG1YThrVall1egInal1aeph1aSmMechap---GlyAlaser1levall1ie	261
Db		71217	GATACCGGAAAAGTGATCCACGTTGACCCTTGCGTGGTAGCCGAATCTTGTCATCACATC	71276
Oy		262	GlnH1eThraenGIyPheValserSerTYLlsh1e1le1ysAsp1aGlnVally1str	281
Db		71277	AAGCAACAAGATACGTTCTTGACCGCGCTTAGCGCAACAACATTAAGTTCTGGTCACCGAG	71368
Oy		282	G1YAsePThrVal1ArGrThrg1yInArGI1ea1aseMeth1ysAnGlnProSeRGIya1a	301
Db		71337	CAATCCACGGTCCAAGAAGGCGAGAGATGCCGAATGGCGCATGCCATGCCAGCCCC	71366
Oy		302	AlAleudheGluPheaRgl1eSerArgAnG1YvalTYrVal1aSpProLeuThrVal1leu	321
Db		71397	GTCGAACCTTACTTGTGAAGTGCAGCCCGCAACGTAACGTGATCGATCGATTCTT	71456
RESULT	23			
LOCUS	U32753	11435 bp	DNA	linear BCT 29-MAY-1998
DEFINITION	Haemophilus influenzae Rd section 68 of 163 of the complete genome.			
ACCESSION	U32753.L42023			
VERSION	U32753.1 GI:1573701			
KEYWORDS				
SOURCE ORGANISM	Haemophilus influenzae Rd. Haemophilus influenzae Rd Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.			
REFERENCE AUTHORS	1 (bases 1 to 11435) Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.R., Kerlavage,A.R., Bule,C.J., Tomb,J., Dougherty,B.A., Merriick,J.M., McKenney,K., Sutton,G.G., Fitzhugh,W., Fields,C.A., Gocayne,J.D., Scott,J.D., Shilly,R., Liu,I.I., Glodek,A., Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Heathlow,E., Cotton,M.D., Uterback,T., Hanna,M.C., Nguyen,D.T., Sauder,D.M., Brandon,R.C., Fine,L.D., Fritchman,J.L., Funtman,L., Georgiagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M., Smith,H.O. and Venter,J.C. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd Science 269 (5223), 496-512 (1995)			
JOURNAL MEDLINE PUBMED	95350630			
REFERENCE AUTHORS	2 (bases 1 to 11435) Tatusov,R.L., Mushagian,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodovsky,M., Rudd,K.E. and Koonin,E.V. Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli Curr. Biol. 6 (3), 279-291 (1996)			
TITLE				
JOURNAL MEDLINE PUBMED	96398784			
REFERENCE AUTHORS	3 (bases 1 to 11435) White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D. Direct Submission Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA 4 (bases 1 to 11435) White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D. Direct Submission Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
REFERENCE TITLE				
JOURNAL				
REMARK				

REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REMARK COMMENT FEATURES	
SOURCE	
corresponding H. influenzae genes 5 (bases 1 to 11435) White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D., Peterson,J., Hickey,E., Dodson,R. and Gwin,M. Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA The whole genome was shifted by 588 nucleotides for a new start on Sep 30, 1996 this sequence version replaced gi:1221405.	
gene	1. .11435 /organism="Haemophilus influenzae Rd" /db_xref="taxon:71421" complement (106. .678) /gene="HI0699" complement (106. .678) /gene="HI0699" /note="similar to GB:I13261 SP:P30856 GB:Z21496 PID:394720 PID:475995 percent identity: 68.25; identified by sequence similarity; putative"
CDS	/codon_start=1 /transl_table=11 /product="peptidyl-prolyl cis-trans isomerase, FkBP-type (slayd)" /protein_id="AAC22358.1" /db_xref="GI:1573702" /translation="MKRKNNVVISISQYRTQDGLVDENAPANPLEYLGGHNLVIG LEKALFAGVEGDGEVAVQPEEGYCAVSEMNVRPVDPFGVDELEVGRPLADTDI GPVPVYTEIDGDEVVDGNHLAGOELHTVEVAAREATLEINAHGHGAHSDD DEHGCGCCGHHHNHRHNSCCGCGR" complement (758. .1219) /gene="HI0700" complement (758. .1219) /gene="HI0700" /note="similar to SP:P37163 GB:U14003 PID:537096 GB:U00096 PID:1790704 percent identity: 54.20; identified by sequence similarity; putative"
gene	/codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC22359.1" /db_xref="GI:1573703" /translation="WSKLAEQLQETRIEEDLINDSEPNALYIIIEHHAHDFDLLE KIANDFKAGEVSEAEPFDGDKPIFCPDIISEVELKAEIIDAOCKELLPLEKEN GIYDGWGYFEDENADDDEXDGEFLDEDEXDGEFFDEDEDESPRVH" 1347. .2347 /gene="HI0701" 1328. .2347 /gene="HI0701" /note="similar to GB:L07942 SP:P36664 PID:1036739 PID:1036738 GB:U00096 percent identity: 44.24; identified by sequence similarity; putative"
CDS	/codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC22360.1" /db_xref="GI:1573704" /translation="MLEOLPYALAKTPPKTALLKACADFIYKEHLGYEMSNGSERV AUYVRTDCNTLFVGSKLAKFAGYSRRMNGYAGLKRRRAVTETGFCLQNPGETPPDS QEFLDGEILITVTRHNKRKTGTSGLEAGYPDILIRGAESSELARLDVFANFGPNP TEORFGRDGNLTQALRWAGBELIKVDORKRSYLSSAAASEIENLVAAIEKSTJNO VLNDIVLOLAGSHSWFKADEKDITLAOVLEMODILLTPLIGEIIILAASEIENEIV NOSADPDIPTKQGRMKAAARRPILMKAKGSWAPEPGRLKFYLPAGSVATALVRELV NYTEE"
gene	2357. .3106 /gene="HI0702" 2357. .3106 /gene="HI0702" /note="similar to GB:U00096 PID:1789101 PID:882637 percent identity: 56.50; identified by sequence similarity; putative"
CDS	/codon_start=1 /transl_table=11


```

/product="stationary-phase survival protein (sure)"
/protein_id="AAC22361.1"
/db_xref="GI:3212202"
/translation="MRILVNSDGFHAGIOLVATELRKIAEVIIVAPDRNRSAASS
LTVLEPRRLHNDGDCVNGTPADCVHLNGLSGOVLVSGIAGNCGMDTTY
SGTLAALBGRHLGDLPAIVSLDROHETPAARVCDLIPKLOHQLNPRRIININP
DLPREELKGVKVCRIQYRSSSVETIKORDPRDETIVIGSALPEDESEGTDFYAVKN
GVYITPIQADLIRHSLSLQNLWDFETK"
3115 3683
/gene="H10703"
3115 3683
/gene="H10703"
3115 3683
/note="similar to GB:L10653 SP:P36684 PID:148929 percent
identity: 72.25; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="lipoprotein B (lppB)"
/protein_id="AAC22362.1"
/db_xref="GI:1573705"
/translation="MKIFGTMYDKTMEWSKRFAVFWLSEVSEIAEFPIDPVLMI
PMSKSKSAVKFAFYALASVIGIIGVATDMVENIVQMGVAHMAKAVSM
PEOMGVLVFVAGSPPIPYKVFILCAGVMQMAFPFVITAFVSRLARFLVAKLAAMG
GEKFAALKRSIEIIGMSVVAVIAVFIKEN"
3702 3887
/gene="H10704"
3702 3887
/gene="H10704"
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=1
/product="H1 influenzae predicted coding region H10704"
/protein_id="AAC22367.1"
/db_xref="GI:1573706"
/translation="MKLSLVLVAGFLVGCASVSDPDALPMGINQVPEGTAVAG
GSFPEIQKNTIPQMK"
3911 5128
/gene="H10706"
3911 5128
/gene="H10706"
/note="similar to GB:L07869 SP:P33648 PID:404099
PID:433185 PID:882635 percent identity: 48.58; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="lipoprotein"
/protein_id="AAC22363.1"
/db_xref="GI:1573707"
/translation="MKKSFLILPLSLVLSACTSNFPAPISDADGNFSPVIOVNGS
NNGAMQPEIQKSLPTTGMMVTPQONFOPINQOPTAPACAPSPPTVVSAT
VOTGVTKTVADCVGDGHNIPRNPNTPNAPYKISKSGSYGNTYKNGKGTMLIAY
LAGIDVKELANLNLSEPNLSIGOVKISCDIKVTTVSVQVSTAIPTAVP
AVTTPGANGTQISDGTITGIRKSEAGTSPSVATSTNOSTVANNANSTNSV
VAPTLNVVMQWPTSGNIIOGFSSTGNGGIDISGRGCAVKAAGRIYVANAAR
GVNGLIILIKNDDELAVAHNDKILVADQGEVKAQDIAKKGSGTNTVTLKHFIRK
GKSDVPRVRLPRH"
5261 7846
/gene="H10707"
5261 7846
/gene="H10707"
/note="similar to GB:M64730 SP:P23909 PID:146906
PID:882626 GB:U00096 percent identity: 70.99; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="DNA mismatch repair protein (mutS)"
/protein_id="AAC22364.1"
/db_xref="GI:1573708"
/translation="MISEENFQHTPMQOYLKLAENPDILLFRMGDFEYELPDA
KKAALALDISLTRKGSAGCPIMAGMRYHVBGLAKLVOLGSSVALICEIGDPTS
KGPVERQIVIVTGVSDLEALPERDNLIAAYOEKKEGLTILMTSGRFLQCEP
ADKETTRAEIQRIPAEVILYCEEFNEMAIHCKGLRRRPIWEEFLSTALITLNRQG

```

```

Alignment Scores:
Pred. No.: 0 00115
Score: 219.50
Percent Similarity: 44.69%
Best Local Similarity: 27.84%
Query Match: 13.14%
DB: 1
Gaps: 10

US-10-018-706-2 (1-322) x U32753 (1-11435)

QY 71 AsnArgTgInValIysGInIYAsPThrValSerIysIleAglInArgTgIleu 90
DB 4343 AATACCTATTAAGTAAGTAACAAAGCGGATACATGTTTAATCGCTTACGGCGGAT 4402
QY 91 AsnTrpArgSILuIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGly 110
DB 4403 GATGTAAAGAAATTTGGCGCGGCTTAATACCTATCCGAACTTAATTAATTAAGTTAGGG 4462
QY 111 GlnTrpLeuThrIleuTSPSerGlyAspLeuIleValArgSILuArgSerIleSerSerGly 130
DB 4463 CAAGTTTAAAGATTCTTAATTTGACATATAAAGCGTTACTCAACAGGTTCT 4516
QY 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
DB 4517 -----GTAACAAACCTCGCAGTCACAGCATCAACAGCAAGCGCGTTAAA 4561
QY 151 GlnHisProAlaValGlnIYsProThrPro 160
DB 4562 -----CGAGAGTCACTTAT--ACCCCGGCGCAACCGTACTCAATAGATCTGAC 4612
QY 161 -----ProValValValIleValIleValSerProThrProThrProProValVal 175
DB 4613 GGATCATTAATTTGGCCCAATTAATCAAGAGCTGGACATCGCTGATGACCTGACGCA 4672
QY 176 GlnGlnProAlaProValAlaProProValThrGlnIleAlaProHealIleThrGlySerSer 195
DB 4673 ACTTCATCAACACAGGTTACATCATCTGTAATTAATGCAATAGCACACCAATTAATTC 4732
QY 196 GILValMetGlnPheArgTyrProValGlyAlaThrAsnProValValArgArgPheGly 215
DB 4733 AATGTCGTTGCG-----CCAAATC--GCCTCAAT--GTTGTGTGGCAATGCGCG 4777
QY 216 ThrAla-----ThrValAlaGlySerThrValIleThr-----SerAsnGlyMetTyr 230
DB 4778 ACTTCAGTAATATCATCAACAGGTTTCTCAAGCAACATGCGGTAAACAAAGGAATGAT 4837
QY 231 PheSerGlyIleArgAspGlyAspLeuIleAsnAlaSerAsnIleGlyThrValIleGlyAla 250
DB 4838 ATTAGTGTTACAGAGACAAAGCTGTAAAGCGGCTCACAGGCGCAATAGGTATGCA 4897
QY 251 AsnHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheVal 268
DB 4898 GGCATGCTTACGTGTGTACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4957
QY 269 SerSerTyrIleHisIleIleValAspAlaGlnValIleThrGlyAspThrValArgThrGly 288
DB 4958 AGTGCTTATGCGCATTAACGAAATTTCTGTGCGCATCAACAGAGCAAGCAAGCGAGT 5017
QY 289 GlnArgIleAlaSerMetIleValAsnGlnProSerGlyAlaAlaIleuPheGlnPheArgIle 308
DB 5018 CAACACATCGCAACAAAGGATGCTGTGTAACAAATCCGTGAACATCCACTTTGAATTT 5077
QY 309 SerArgAsnGlyValTyrValAspProLeuThrValIleu 321
DB 5078 CGCTATAAGGTAAATCATGAGATCCAGTACGTTACCTA 5116

```


RESULT 24
LOCUS AE011804 11555 bp DNA linear BCT 29-MAY-2002
DEFINITION Xanthomonas axonopodis pv. citri str. 306, section 182 of 469 of the complete genome.
ACCESSION AE011804 AE008923
VERSION AE011804.1 GI:21107911
KEYWORDS Xanthomonas axonopodis pv. citri str. 306.
SOURCE Xanthomonas axonopodis pv. citri str. 306.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas.
REFERENCE 1 (bases 1 to 11555)
AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J., Chambergo, F., Chapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Medeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.R., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sana, J.A.D., Silva, C., de Souza, R.F., Spínola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, F.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.
TITLE Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
JOURNAL Nature 417 (6887), 459-463 (2002)
MEDLINE 22022145
PUBMED 12024217
REFERENCE 2 (bases 1 to 11555)
AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J., Chambergo, F., Chapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Medeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.R., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sana, J.A.D., Silva, C., de Souza, R.F., Spínola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, F.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-2001) Departamento de Biologia, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil
FEATURES
source location/Qualifiers
1. 11555
/organism="Xanthomonas axonopodis pv. citri str. 306"
/strain="306"
/db_xref="taxon:190486"
/note="pathovar: citri"
186. 1850
/gene="pyrG"
/note="XAC1716"
186. 1850
/gene="pyrG"
/note="XAC1716"
186. 1850
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="CTP synthetase"
/protein_id="AAM36583.1"

gene
CDS
/db_xref="GI:21107912"
/translation="MTPLIFVTGGVSSLGKGIYAAASLAEILAEGLKVTMMKLDPLYI
NVDGTSPPHGVVYVTDGAEITDLGLGKRIYVTRLSKNSVTRGTYENYTRK
RGDYLGAQYQVPHITDEIRICIDEADNAGEADYIIGTGVGIBELPKEAQR
TERGAECAMPHMLTIVYIAAGSLKTRPTPOHSVKEIRISTIQDVLICRSEQAVPS
ERRKIALFTVSEBAVISCPIIDVLYCMPLELLKQGLDELYIDQFKLRKVAADLS
MAAVADVAKHLDENVNAVGVKVDHDAKSVAEARHGGRLQRITKNLAEADL
EGSDMSALODIDGILVPGFGDFGEGVSKPAREQKYVFGICGMQAAVDYAR
HVADLDANSTENDROSHPHVLGITEMRTATGVEKREKSDGLGTMRLGLOSRLK
RGTLAERVGVKDVAVERRHRHYPENNRYTQLEDPAGIVSGKSMDDLVVEVLPDRT
HPWFLACAPHEFLSTPRDHPFLFGVRAAREKAGKLUKERA"
2202. 3032
/gene="kdsA"
/note="XAC1717"
2202. 3032
/gene="kdsA"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="2-dehydro-3-deoxyphosphoocronate aldolase"
/protein_id="AAM36584.1"
/db_xref="GI:21107913"
/translation="MKLCDFEVLGDQPLFIAGECVIESNQQLQDVAGKLEITGKLG
VNFIFKSPFKARSTSGTSFRGLBGLKVLDAVKQIGVPLTVDHAYPNMVA
VVDVLOTAPLVQITDIFKAVCAAGKPVNIKGGFLAPMMPKPVDAKSTGNOIV
CBGASFGYNNIVSDMSLSVMRDTGCPVVDATHSYQLPGGSGSSGGREPVLA
RAAVAVGISGLPAETHRDPKSLSDGNAMPIDMEELLLETIMELDTVTKHGARRA
"
3138. 3389
/gene="XAC1718"
3138. 3389
/gene="XAC1718"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAM36585.1"
/db_xref="GI:21107914"
/translation="MAVMATLITRSTGPPGDDGNAMQTRLPKRTITLAVMAAVGSL
ATGSAALRIKQPEVALAULTLVLVLLSLGLGMA"
3514. 4806
/gene="eno"
/note="XAC1719"
3514. 4806
/gene="eno"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="enolase"
/protein_id="AAM36586.1"
/db_xref="GI:21107915"
/translation="MTTIATILAREILDSRGNPTLEAVTLDDSGFRAAVPSGASTG
TKRAVELRDGDXTIRKGVRAVDNNGVIAETLKPKADQGLDLDLGTEN
KGRGANALVGLSAAHAAVAASKQMLQVLSITADVAALPVPMNITINGADAN
NDROEFMTLVGCGSFSFSLRAGTEIFSLKSYLKQHGSLTAGDGGGAPDPRSRY
BALDTIEATGKAGYTRAGEDILGSLDVASSEFDNGKYNLVGENKRLTSBQFVPLAD
WVAQYPLISIEDGAEEDWAGWKLTDRVERKQVLDVDDLFVNPXKLFKQSIDGTAN
AIIKVNQIGTLTETLEIAIMAHANAYASIVSHRSGETETTTADIATATATQIKG
SLCSGDVAKYNQILRIEALSGARVAGDAPVSIRK"
4810. 5175
/gene="XAC1720"
4810. 5175
/gene="XAC1720"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAM36587.1"
/db_xref="GI:21107916"
/translation="MRWRWLLLVLAVLAMLVQYRFWFGGNSGSEVMLEAQAHQIQ

gene
DNEGLRQNOALAEVNDLKGFAIEERARSELGMIKPGTEFVVEDAPLPAPASA
BASAPAOQAEPEPDVDPHP"
5172..5972
/gene="ispd"
/note="XAC1721"
5172..5972
/gene="ispd"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="4-diphosphocytidylyl-2C-methyl-D-erythritol
synthase"
/protein_id="AA036588.1"
/db_xref="GI:21107917"
/translation="MTGSIVAIYPAAGRGTRPGGPIPKOYIAGGPIAMATIAALAA
HPALAGMVAIAPGADMPGTAVOSRVTCTCGATPAASVSGILAPESVADPF
VVAHAKPRLALADRLLEIGSGDVGIIAIPVDTTKRAGDGTGTEPREET
WRALTEPLRRHQLIKGILASAGVTVTBAMMERIGARPLIVEAEDNFKTTTFA
DLARFELLRNDGPGTRDPESAHPOSVASAFSGSARAPETI"
5976..6476
/gene="ispd"
/note="XAC1722"
5976..6476
/gene="ispd"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="2C-methyl-D-erythritol 2,4-cyclodiphosphate
synthase"
/protein_id="AA036589.1"
/db_xref="GI:21107918"
/translation="MSFNFRIGQGYDVHAFPGDDVWLGVRVAHSHGVASHDGDV
LHALCDMLGALADIGRHPEDDARKDADSQIOLHCGILREGRMGNDITV
ICERPKYGPAAVARRIADILAIELDAVSATTSBOLGFTGRGEGIAAOAVILGR
IAAPHQ"
6589..7698
/gene="XAC1723"
6589..7698
/gene="XAC1723"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="hydrogenase subunit"
/protein_id="AA036590.1"
/db_xref="GI:21107919"
/translation="MSERSILLPRAHGAIVISAAMRSTPDDDELPAFSPSGEHL
LITRKQNTYVIAKILAHAGIAPKGVSAKIKDPAVTTORPSYHLPRIADPA
ALDTOMOVSSMNRKLOKALHGNRPVITLBOYGERALHIOAIAAPILPW
FGEORFGDGNVAAIANRFAHVDDSTILPAPSRRIADHDSMTLSAARILPW
RVLARVAGSSMDGLSEAWMDSSRSVFEPERSLARLARDIHPGPIKMGAG
QUNSTDAAVAEGALSDPSTIALRQGLRAGLKERALRLRPHGLTYQLEPQTLQ
LAPLAPPECAIVAMELDVADAGRFNVADRAAD"
complement(7951..8490)
/gene="XAC1724"

gene
Alignment Scores:
Pred. No.: 0.0016 Length: 11555
Score: 219.50 Matches: 81
Percent Similarity: 40.07% Conservative: 38
Best Local Similarity: 27.27% Mismatches: 111
Query Match: 13.14% Indels: 67
DB: 1 Gaps: 11

US-10-018-706-2 (1-322) x AE011804 (1-11555)
Oy 30 Leu1aag1yCysalaser1yProthry1yAenSer1yTherSer1ySerg1ySer1ySarg 49
Db 10736 CTGACCGGCTGTAGAGCGCAACGCTGTGCGTTCGCCCGGAGCGAGCGGTAGCGGGCG 10795
Oy 50 TherSerg1yGlyLeu1a1le1leg1ySerg1yVal1le1ThraSerg1yGlyVal 69

Db 10796 TCCTCCGCTCCAGTCC-----CGCCCGCGCCCTTCGCTG 10831
Oy 70 ProAsn-----ArgTyrGlnValLysGlnGlyAspThrValSer1yIleAlaGln 86
Db 10832 CCGCGTCCGGATGCCACCGGTACCGAGCGCGGATACCCGTGAACGGAATCTCGCG 10891
Oy 87 ArgTyrGlyLeu1a1le1n1yParGlu1le1Gly1yIleAsnAsnLeu1a1n1Ser1yThr 106
Db 10892 CGTACCAATATCAACCGCGGATCTGGCCGCTGGAAATGCTGCTTCGCGCAACACC 10951
Oy 107 11TyrThr1yGln1yPhe1yLeu1a1le1n1ySer1yAspLeu1yValArg1yLysSer 126
Db 10952 ATCTATCCGCTCAGACGCTCAAGTTGTAT----- 10981
Oy 127 11SerSerg1yValAla1n1yLeu1a1le1n1yProSerProValAlaValGlnSerSarg 146
Db 10982 ---CCGCGCGGTCCAGCAACCGCGGAGCGAGCGCGCGCTGGCTGGCGGACATGTGCG 11038
Oy 147 ProProValGlnGln1yProAlaValGln1yProThrProProValAlaValLys 166
Db 11039 CCGCG-----CGCCCTGCTG----- 11056
Oy 167 LysProThrProThrProProValAlaGlnGlnProAlaProValAlaProProValThr 186
Db 11057 -----CGATCGCGGACCGGTC--- 11074
Oy 187 GluAlaProPheAlaThrGlySerSerg1yValMetGlnPheArgTyrProValGlyAla 206
Db 11075 -----AGTAGCGGA---TTCTCTGGCGGTGGCG-----GCC 11104
Oy 207 ThrAsnProValAlaAlaArgPheGlyThrAlaThrValAlaGlySerThrValThrSer 226
Db 11105 GATGCGCTGTGTGGTGGGACCTC-----GTCACTGGCGAAACC---ACCAAG 11149
Oy 227 AsnGlyMetTyrPheSerg1yArgAspGlyAspLeu1leAsnAlaSerAsnAlaGlyThr 246
Db 11150 CAGGCGCTAGACATCGCGGACCAAGCGGCGGACGCTACGTGTGCGCGGATGGCGTG 11209
Oy 247 Val1leGln1a1aPhe1a1aSerMetAspGlyAlaSer-----11Val1leGln1a1aThr 264
Db 11210 GTGTGTATTCGGCGCGCGGCTGTGTGGTACCGCGCACTGAATCATCATCAACACAAAC 11269
Oy 265 AsnGlyPheValSerSerg1yIle1yIle1yAspAlaGlnValLysThrGlyAspThr 284
Db 11270 GACCAAGCTATTCGCTATCGCATACCGGTACACGAGCGCTGTCAACGAGGCGCAAGC 11329
Oy 285 ValArgThrGlyGlnArg1yIleAlaSerMetLysAsnGlnProSerg1yAlaAlaLeuPhe 304
Db 11330 GTCAACAGGGGCGAGCAAAATGGCGAAATGGCGGCGGCGGCTCCCGGACATGTTG 11389
Oy 305 GluPheArg1yLeu1a1le1n1yValLysProLeu1a1le1n1yValLysProLeu1a1le1n1y 321
Db 11390 CACTTCGAGTCCCTCAACGCGAAGCGGCGGATGATCGTTGCTGTATCTG 11440

RESULT 25
AE002061 14193 bp DNA linear BCT 22-NOV-1999
LOCUS DEINOCOCCUS radiodurans RI section 198 of 229 of the complete
DEFINITION chromosome 1
ACCESSION AE002061 AE000513
VERSION AE002061.1 GI:6460095
KEYWORDS
SOURCE Deinococcus radiodurans.
ORGANISM Deinococcus radiodurans.
Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
Deinococcus radiodurans.
1 (bases 1 to 14193)
REFERENCE
AUTHORS White,O., Eiken,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
Moffat,K.S., Qiu,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
Vamathevan,J.J., Lam,P., McDonald,L., Uetzerback,T., Zaleski,C.,
Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.

TITLE Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1
JOURNAL Science 286 (5444), 1571-1577 (1999)
MEDLINE 20036896
PUBMED 10567266
REFERENCE 2 (bases 1 to 14193)
AUTHORS White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J., Haft, J. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L., Moffat, K. S., Olsen, G. J., Jiang, L., Paulsen, O., Karp, P. A., Smith, T. F., Vamathevan, J. J., Lam, P., McDonald, L. L., Utterback, T., Zalewski, C., Makarova, K. S., Aravind, L., Daly, M. J., Minton, K. W., Fleischmann, R. D., Ketchum, K. A., Nelson, K. E., Salzberg, S., Smith, H. O., Venter, J. C., and Fraser, C. M.
TITLE Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
JOURNAL
FEATURES
SOURCE Location/Qualifiers
1. 14193
/organism="Deinococcus radiodurans"
/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
/gene="DR2286"
/complement (90. 1598)
/gene="DR2286"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA011844.1"
/db_xref="GI:6460106"
/translation="MCCDNVPRSPHSMTARVNPTRITLRSCTIEGSMRLKFNRA SPESGPGVDFVDDKGEKRPVAVDRTQWQVGLGELYDHLGVNDUTLTPTRGRGAE VETVYKRYSTPRARQTRQKPPSSGRTETTRKEBGESEAROGASPSVREYRQQA GSILGRTAPSGGAINSLGLGEERRPAGGEEGPPSLKPIAFAFTQVQPERQELT LPSLASFTGRSREABVPAPGAASGSPDPVALSPVSPVPAVQEVSDAPAR AQGALGEOIVPERLASGRLTLGRLVRLTADLGPAYGVTLATDRALSSPERD SGDDHRLITTEEDRARGVRLTREALVSLIEHARLAPLGLDLRGYWRAGDVLDISA ASVAELVSALHARCTFTYLLTISQAPASLVNVSLEAQLNGVSTDDMNLIDTL ARPPLAVTLPENQPLNRSTVLDLADLVAGVRRRLARPGGTGGRG"
/gene="DR2287"
/complement (1808. 2098)
/note="similar to GB:AB000782 percent identity: 64.18; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, AsnC family"
/protein_id="AA011844.1"
/db_xref="GI:6460096"
/translation="MSGMTAIVVQQAERHRIQETAALAAVSVREVSVTGEMDIV AIVKALEYDLDDVVTGHLKRIKIDIMRTQMLAFRTYNEVLQGFGLDG"
/complement (2232. 3733)
/gene="DR2288"
/complement (2232. 3733)
/gene="DR2288"
/product="Ribosomal RNA"
/complement (4055. 4567)
/gene="DR2288"
/complement (4055. 4567)
/gene="DR2288"
/note="similar to GB:U00096 PID:1787283 percent identity: 64.29; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA011845.1"
/db_xref="GI:6460097"
/translation="MPDELVQGDIAHQPDVAVTANKQLMGGGGVGDIYHRAQPERL LQAIRPGTPTGTAITPAFDLERQGVKVIHAVGPIWRGQHGELLAGVRESL RLGVNCGRSVAFPSISTGVYGYPLDRAADIALATIDFLRSHDLSVRVNLGADAL

gene
CDS
HFERALAQ.
/complement (4567. 5385)
/gene="DR2289"
/complement (4567. 5385)
/gene="DR2289"
/note="similar to PID:1651722 percent identity: 68.11; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="guanylate kinase"
/protein_id="AA011846.1"
/db_xref="GI:6460098"
/translation="MGESQAPGAARSGTGDARTDDEAVSVASACMASAPDRPA GTPQGRGLIIVTGASGVGKTLREMLAGQVDFSTWTRPREPGVNGDYVY SPAEPLAKQONGFLEHAQVGNHGYPIPIAALERGQDVLEIEVEGAMQKDM GEOAIIYTPSPSTELRLRLRGATETPERIKRLTRADEYQAADPRVYVNDL DRAVELAVQQAERAAQKAERHTPEEQARALADVASTALREALQOVES"
/complement (5351. 6193)
/gene="DR2290"
/complement (5351. 6193)
/gene="DR2290"
/note="similar to GB:A009126 percent identity: 55.74; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="pseudouridylylate synthase I"
/protein_id="AA011847.1"
/db_xref="GI:6460099"
/translation="MDADPRLYAPPGFTRLRLTVAMGDRDYAGQORNAVQDET LQALHAGGEGALRPVSAGRTDAGVAEMPLHMDVPTFRVPLVOLPRLANLPA SVSLAEVAPAGFAHAFSCTERRVYRLVAVPQRLHAGRALHYGPIDAEAMNRA ASSLIGHDPAFATREDROTVDLRLLEVQPGELMDIVAGESFRRHVRGLVGTLLVNGOKLGAEEVGLASERRARAGANVPGGLYFAGAYGRVREGRGNGEPEPT HRRR"
/gene="DR2291"
/complement (6309. 7673)
/gene="DR2291"
/note="similar to PID:1653709 percent identity: 59.04; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell wall glycol-glycine endopeptidase, putative"
/protein_id="AA011848.1"
/db_xref="GI:6460100"
/translation="MLANVPLSLIRLOMKSSPPAPVYLGVTGILLAGALLGAGAE SYRVKGETLYRITALNAGLSEETVQOANPYLRGHALYAGQMLTIPKPLPPTFRYR KGBDLKKAORLGVSGDIRDNPQIDRSLNAGVRLRPARLIAORAAEOAAQOK AAOQKAAOQAAQAAQAAKRAAQAALAAKPPPTTHRVVEIGTFYSVARVGIN PTLQGVNPRILAQTLNAGVSLVAPRLAPVAPARRAPVTLVAPRAPASPTPAP TAPAPAPAPAPAPAPAPAPAPVQAQTPBHAAVAVQVTSLSIMQPLPQYGI TSDFGWRVLDGEREKQGLIDVAPPGTPTVAAASGVTOAHLDETYGMQWTVYIQHPD GMQTRVAHLSRISVEAGQIVROGERVAGVSTGVGPHLFGLYRWMDPHNPLAFYA GGLD"
/gene="DR2292"
/complement (7675. 7932)
/gene="DR2292"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA011849.1"
/db_xref="GI:6460107"
/translation="METAPHTWPSVTVVYELNGKAEANTRLQWLEQLPQAGFAGAE LIDSPPAQGLIAVASHMSGPLPRLTPPAGAKHWTFRVLAQR"
/gene="DR2293"
/complement (7968. 8956)
/gene="DR2293"
/note="similar to GB:D12650 SP:P30865 PID:1208970 PID:216647 GB:U00096 percent identity: 45.69; identified

by sequence similarity; putative"

/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA011839.1"
/db_xref="GI:6460101"
/translation="MGKMRPFLPFLPSSSLVNGMFGVALGMAAGVSGRTPLTL
LVALPPLWLPALPGLALMSVGRGRLGLTLGLAMAGLHMRPOADGELRVL
YNNARSGQGTAAALSTLQASDADLILOETNLIDPARKALRGLSGVALTGREAV
TLTNPVLRQTOHAVPSTRTFAETWKGDPGRTRLRVNNAHLGVLITTSAGTAV
RTRAQARQVALLCRIAHPPGVLVGLDLNTPRGRRLRRLTGCGPAAHRAAGG
CGWTFGLRLRLHLLVRDLRPTLRVLSMAGSDHRLDLVEYRMR"
/complement (8987..9295)
/gene="DR2294"
/complement (8987..9295)
/note="similar to GB:AE000657 percent identity: 71.43;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="periplasmic divalent cation tolerance protein"
/protein_id="AA011840.1"
/db_xref="GI:6460102"
/translation="MSLVLVLTLPERRAOLARTLTERLAGCVNIIPIQISTYRMWG
EVAEDPSILLIKTVGEQYPALEMRITSLHPYEPVETVALPFDASPEFOSWLDSTG
"

gene
CDS
complement (9342..9755)
/gene="DR2295"
complement (9342..9755)
/gene="DR2295"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA011846.1"
/db_xref="GI:6460108"
/translation="MLCSAPMDASPSAPSAHRTWRLRSLRQLRLPLTWMTCL
LLLSGVQAAFOIGNNVRSWMTWQTEQVGRVALQSDILMKDAEAAADPAVILQ
"

Alignment Scores:

Pred. No.: 0.00146 Length: 14193
Score: 219.50 Matches: 73
Percent Similarity: 43.93% Conservative: 50
Best Local Similarity: 26.07% Mismatches: 116
Query Match: 13.14% Indels: 41
DB: Gaps: 9

US-10-018-706-2 (1-322) x AE002061 (1-14193)

QY 60 SerGlnValIleThrAspSerGlnGlyValAlaProAsnArgTyrGlnValIleGlnGlyAsp 79
DB 6667 GCACNACCTCGCGCCGCAAAACCAACCCACCCACCGGTGCAATCGGCAC 6926
QY 80 ThrValSerIysIleAlaGlnArgTyrGlyLeuAsnTrpArgGlnIleGlyHisIleAsn 99
DB 6927 ACCTTCTACAGCGTGGCGGGGCGGACATCAACCCCATCGCCCTACAGAGTCAAC 6986
QY 100 AsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuTrpLeuTrpSerGlyAsp 119
DB 6987 CCGCGCTCGACAGGCAACGCTGACGTGGCGGCGGTGAGCTGAGCTGCTGCTCCCG 7046
QY 120 LeuIysValArgGlnArgSerIleSerSerGlyValAsnThrAlaHisThrProSerPro 139
DB 7047 CTACGT-----CCGCGACCT 7061
QY 140 ValAlaValGlnSerSerArgPro---ProValGln-----GlnHisProAlaVal 155
DB 7062 GTGGCC-----CCAGCCCGCCCGCGCGGTCTCGTGGTGGCCCGACCCGACCT 7115
QY 156 GlnIysProThrProProValValValValIleIleIleIleIleIleIleIleIleIle 175
DB 7116 GCATCACCACCGCCGCGACCTACCGACCCCGCGCCCGCCCTGCGCGACCTCTCTCT 7175

QY 176 GlnGlnProAlaProValAlaProValThrGlnAlaProPheAlaThr----- 192
DB 7176 GCGGCGCTTCTCTCTGACAGCTCTGCTCCCAAGAGTGGCGGACGAGCCGCAACAT 7235
QY 193 -----GlySerSerGlyValMetGlnPheArgTyrProValGly 205
DB 7236 CATTCCCGCGTGTGTGGCGCAACACGACGACCAAGCTGTGGAGTGGCGCTGGCCG 7295
QY 206 AlaThrAsnProValValAlaArgIlePheGlyThrAlaThrValAlaGlySerThrValThr 225
DB 7296 GGCTACGAGCAGCTACCGACGAGCTTGGTGGCGGTGTGGACGCGCAACGCAAAA 7355
QY 226 SerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnIleSerAsnAlaGly 245
DB 7356 CACCAGGCGATCGACGTAGGCGCCCTCCCGACCCCGGTCAATCGGCGCGCAGCGGG 7415
QY 246 ThrValIleGlnAlaAspHisAsnMetAsp-----GlyAlaSerIleVal 260
DB 7416 CCGGTCAATCAGCG-----CACCTCGACGAGACCTACGCTGGGCTGTGACAGTAGTA 7469
QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleIleAspAlaGlnValIys 280
DB 7470 ATTACAGACCCGACGCGCTGGACAGCCGCTACCGCACCTCAGCGGTATCTCGTGGAG 7529
QY 281 ThrGlyAspThrValAlaArgTyrGlnArgIleAlaSerMet-----LysAsnGlnPro 298
DB 7530 GCGGCGCACTCGTGGCGCAGGCGGCGGCGGTGGGCGACGACGCGCGCGTGC 7589
QY 299 SerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeu 318
DB 7590 ACCGCGCCGACCTG---CACTTCGGGCTTACCGCAACCTGGACCTCAACACCGCTG 7646

RESULT 26
AE013686 10592 bp DNA linear BCT 26-JUL-2002
LOCUS
DEFINITION
VERSION
ACCESSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCES
AUTHORS
TITLE
JOURNAL
PUBMED
FEATURES
SOURCE
gene
CDS
modification: Lipoprotein"

1 (bases 1 to 10592)
Deng, W., Burland, V., Plunkett, G., III, Boutin, A., Mayhew, G. F.,
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,
Blattner, F. R. and Perry, R. D.
Genome Sequence of *Yersinia pestis* KIM
J. Bacteriol. 184 (16), 4601-4611 (2002)
2 (bases 1 to 10592)
Deng, W., Burland, V., Plunkett, G., III, Boutin, A., Mayhew, G. F.,
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,
Blattner, F. R. and Perry, R. D.
Direct Submission
Submitted (21-FEB-2002) GenBank, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..10592
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
83..1084
/gene="nlpD"
/note="y0833"
83..1084
/gene="nlpD"
/functions="membrane; macromolecule synthesis;
modification: Lipoprotein"

```

/note="residues 1 to 333 of 333 are 58.65 pct identical to
residues 1 to 379 of 379 from E. coli K12 : B2742;
residues 1 to 333 of 333 are 58.65 pct identical to
residues 1 to 379 of 379 from Genpept :
>gb|AA057849.1|AB005502.3 (AB005502) lipoprotein
[Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=1
/product="lipoprotein"
/protein_id="AA064418.1"
/db_xref="GI:21957563"
/translaction="WSTGSEWIRLRVAACTVSLVLMVCTNDNSTAPISVGDGRS
GTMLSKANTDSSGGRIYVRSYDNI PKGSYGNATYVKGDTLFIYAMITGNDPRDLA
AKNIAPYSLAVGKQSIQIOLGNGSGGMLAADATSSGIAQSPNIOQTITIOQTITV
DOSTASVSGNSGKQVQKMLPSSGAVVATPATPSSSVSPASNGPVGSKWMP
DKTIDSPSASEGNGKIDGASRGOPILATASGRVYAGNALGYNGLIITIKNDY
LSAYAHNDMLVRCQEVKAGQKATMGSTSTSVRLHFEIRYKGSVNPRLRYLPOR"
1139. .2137
/gene="rpos"
/note="y0834"
1139. .2137
/gene="rpos"
/function="enzyme; global regulatory functions"
/note="synthesis of many growth phase related proteins;
residues 1 to 332 of 332 are 92.46 pct identical to
residues 1 to 330 of 330 from E. coli K12 : B2741;
residues 1 to 332 of 332 are 97.28 pct identical to
residues 1 to 331 of 331 from Genpept : >gb|AA063391.1|
(U0152) Rpos [Yersinia enterocolitica]"
/codon_start=1
/transl_table=1
/product="RNA polymerase, sigma S (sigma38) factor"
/protein_id="AA064419.1"
/db_xref="GI:21957564"
/translaction="WSSSTLKNVLEHEDADPDENSTETETIDEKALVDEPTESELAD
DELAQVOTRVLDATOLYLGEIGSPLTAEVEYFARRALRGDVSSRRMISNLR
LVYIARYSNRGLADLIDIEGNGLIRAVEKEDPGRFSTYATWIRQITIRAI
MNQRTIRLPIHIVELNVLYRTARSELKSHAEIAEOLDKPVDYRMLRLN
ERISVDPPLGGDEKALDILSDENENGPEDTODDMKOSIVKMLPELAKREVL
ARRGLGAEYAAITVEDGREIGLTRERKROQVEGRRLREILQGISIALRRE"
complement(2497. .5052)
/gene="muts"
/note="y0835"
complement(2497. .5052)
/gene="muts"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/note="residues 5 to 851 of 851 are 84.09 pct identical to
residues 5 to 853 of 853 from E. coli K12 : B2733;
residues 5 to 851 of 851 are 84.09 pct identical to
residues 5 to 853 of 853 from Genpept :
>gb|AA057842.1|AB005501.1 (AB005501) methyl-directed
mismatch repair [Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=1
/product="methyl-directed mismatch repair protein"
/protein_id="AA064420.1"
/db_xref="GI:21957565"
/translaction="WKNNDKUDSHTPMGOYLRLKAQPEILLFRMGDEYELRYSDA
KRAQOLDISLTKRGASAGEPIPAAGVPHSIENTLAVOLGSSAAICEDIGPAT
KGPYERKVRIVTPGTISDEBALQROQNLAAIWOVDAKGFYATLDSGFRVAT
ADLEMAELORTNPDAELPENPEPMSLIEHRGLRRRPLMEFELDTAKOOLNIOFG
TRDILIGVEBOHIALRAAGCLLOVVKOTRSLPHIRGLTMRQOGGIIINDATRRN
LELTONLSGSENTLAILDSCVTFMGRMKRMIMRIRIVITDQOALIGLODI
LAEOQPLRQVQDLERITLRLAKTARRRDLARMKHARQOLPEIRLQPIDVHVN
LALVQGFDELQDLLEERAVETPPVLVADGVVAGSNAELDEWMLADGATVLDL
EIEREKGLDGLTKVGFNGVGYIOWRSQSHVPIHYVRQTLKNAERYIIELEKE
YEDVKLTSKGAALIEKGLYEIEFDLPLPELOLSANALAEIDLNAERLSTN
YSCPTLSDPKGIKIMGRRHVPVEQVAKKEPTISNPLTSSPORMLITGPNMGKSTW
ROTALIVLAILGSGVPADQATIGPIDITFRVGAADLASGRSMTEVETPNTIILH
NATEOSLVLMDIEIGKISTYDGLSLAWCAELASRIKAMTLFATHIFELTLPKME
GAVVNHLDALHGETIAFMHSVQEGNAKSYGLAAALAGVPRVITKARQKLELES
LSNNAASTIDSGQWTLNEBEIPRAVEALEALDPDLSLPROALEWIVYRLKNMV"

```

```

5549. .5683
/gene="y0836"
5549. .5683
/gene="y0836"
/function="unknown"
/note="residues 1 to 42 of 44 are 39.53 pct identical to
residues 84 to 126 of 698 from Genpept :
>gb|AF28325.1|AF204952_3 (AF204951) Esv-1-144 [Ectocarpus
siliculosus virus]"
/codon_start=1
/transl_table=1
/product="hypothetical"
/protein_id="AA064421.1"
/db_xref="GI:21957566"
/translaction="WAGMKLQELQIRNEGHCPASELKDEKORLLETSSSDCRGVG"
complement(5846. .6301)
/gene="rpiB"
/note="y0837"
complement(5846. .6301)
/gene="rpiB"
/function="enzyme; central intermediary metabolism;
Non-oxidative branch, pentose pathway"
/note="residues 1 to 147 of 151 are 44.21 pct identical to
residues 1 to 146 of 149 from E. coli K12 : B4090;
residues 2 to 150 of 151 are 54.36 pct identical to
residues 1 to 148 of 151 from Genpept : >gb|AA063729.1|
(AB06201) RpiA [Pasteurella multocida]"
/codon_start=1
/transl_table=1
/product="ribose 5-phosphate isomerase B"
/protein_id="AA064422.1"
/db_xref="GI:21957567"
/translaction="WLSIATGCDMAIEMKQVIFLOSKNITVVDVSCDQOERTW
PDIAQVLAIAKAGHERGILTCGTGMSISAKVGVRAQCHDIFSACRARSNN
AQIICLARVGPPLATLLEAMDAEFDDGGSTEXERIRRYEQSLHD"
6751. .7899
/gene="y0838"
6751. .7899
/gene="y0838"
/function="putative"
/note="similar to 2,3-butanediol dehydrogenase; residues
30 to 375 of 382 are 34.00 pct identical to residues 1 to
345 of 349 from Genpept : >dbj|BA060856.1| (AB063194)
acetylacetoin reductase [Bacillus cereus]"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AA064423.1"
/db_xref="GI:21957568"
/translaction="MCDKPKKESMSHENTKGNOMAGEIPETMRVAVVAGPROYRFE
RVYPTIDAKEIILVKBGCGICACADTYAFBEABEPFKDDKRPATIKAMITGHEIGH
LVGKGAIVEGELDVRVTEQIVPCWQGRFENHQYVWCHEGDIIVGFSNNGVALEY
LKEPTKAINVHPHPLPEKAILIEPYACAFHARQAKIKIDVAVVAGSGTLGIMI
GAIRKSGAAKLIVDLSDDDLALAKKGCADVLVNSDDVNTLLKEMTGYGCDIYID
ATGOKXAVEOGLTLRLKLGFEVRSVFCDPYVMSIISDRKELDVLSHIGPYCFPL
VIEGISNGDLPTBEVTHIILPEKYAGSIELVKRGVDSLKVLNPNLL"
7963. .8733
/gene="y0839"
7963. .8733
/gene="y0839"
/function="putative"
/note="residues 5 to 256 of 256 are 54.76 pct identical to
residues 3 to 254 of 254 from Genpept : >emb|CAC95595.1|
(AU596164) similar to dehydrogenase/reductase [Listeria
innocua]"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AA064424.1"
/db_xref="GI:21957569"
/translaction="WPNDYKWFDTAFSLSGKVAALITGANGTIGHAIABLYAKARV
LMDCADNVABIEABOLDRDNAGLHCDVSDSOSVQAQVAQAIAGFGOLDILVNSAGIAA"

```



```

shikimate kinase AroL or AroM SW:AROL_ERCHC (P10860) (173
aa) fasta scores: E(): 0, 63.7% id in 168 aa"
/codon_start=1
/transl_table=11
/product="shikimate kinase II"
/protein_id="CAC92450.1"
/db_xref="GI:15981151"
/db_xref="SPTREMBL:Q8ZC15"
/translation="MTQTFVNGARGAGKTTIGKALQALGYRFVTDLDLEMQGTSQNT
VAEVESEBMDGRLRESMALQAVTAPRTVATGGAVALSENAPFMDHGRVYLLAA
SAVALVRLAEEDPEERAPSLTGKPIVEEIIIDLVALREALYQDVAAHVLDQTPSLV
VQIILQMLTGSEMWK"
/complement(209..709)
/gene="YPO3215"
/feature="misc_feature"
/feature_note="Pfam match to entry PF01202 SKI, Shikimate kinase,
score 245.00, E-value 1.1e-69"
/complement(470..544)
/gene="YPO3215"
/feature="misc_feature"
/feature_note="Shikimate kinase signature."
/complement(668..691)
/gene="YPO3215"
/feature="misc_feature"
/feature_note="P800017 ATP/GTP-binding site motif A (P-loop)."
/complement(962..1519)
/gene="YPO3216"
/complement(962..1519)
/gene="YPO3216"
/complement(962..1519)
/gene="YPO3216"
/feature="misc_feature"
/feature_note="Similar to Rhizobium meliloti phospholipid
N-methyltransferase PmtA TR:AA610237 (EMBL:AF201699) (200
aa) fasta scores: E(): 1.1e-07, 25.9% id in 185 aa, and to
Bradyrhizobium japonicum PmtA protein TR:Q9ICT2
(EMBL:Y09633) (199 aa) fasta scores: E(): 0.0057, 24.2% id
in 178 aa, and to Klebsiella pneumoniae hypothetical 22.7
kDa protein TR:O87755 (EMBL:A011907) (196 aa) fasta
scores: E(): 0, 52.9% id in 191 aa"
/codon_start=1
/transl_table=11
/product="putative methyltransferase"
/protein_id="CAC92451.1"
/db_xref="GI:15981152"
/db_xref="SPTREMBL:Q8ZC14"
/translation="MLNIVRNRFKSFQSLQRFASPRVGSILVSPSPMLCOALNIOI
DMTONDAIEAGADGVTKRLTSMNSNGIQAIEQHPFVHLHOLKSRLOIADRP
SBAQLDNRGIDVFCCLPLISTPTKTSITILQOTQRLRARGVLLFYSHLSSELSLS
RIFNMKKIRVAVNPFPAIVYICHPN"
/1787..2002
/gene="YPO3217"
/1787..2002
/gene="YPO3217"
/feature="no significant database hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC92452.1"
/db_xref="GI:15981153"
/db_xref="SPTREMBL:Q8ZC13"
/translation="MIGTDGVGVANVIAETNDRGRSISCHFFIGSPFAHPLANDV
IKNIASPIVLSTIVACSIKKAAXKRLS"
/complement(2249..2770)
/gene="YPO3218"
/complement(2249..2770)
/gene="YPO3218"
/feature="misc_feature"
/feature_note="Similar to Escherichia coli hypothetical protein
YafE SW:YAFB_ECOLI (P30866) (207 aa) fasta scores: E(): 0,
65.1% id in 169 aa, and to Pseudomonas aeruginosa YafE or
PA3119 TR:O87011 (EMBL:U03274) (187 aa) fasta scores: E():
0, 55.2% id in 172 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC92453.1"
/db_xref="GI:15981154"
/db_xref="SPTREMBL:Q8ZC12"
/translation="WQAAADKKLTNIEIQGLAESLPFADOSPDIIVSRYSAAHHMDV

```

gene

KGALAEVREYLRPGKGVLPMDVVSQGHFVLDTIIVTQTVELBDSHVNRNYSAGEMLSTIS
TGSLAVINERVTSDPLRYLETFSSWVAWRMRPAPFAIRAIREFQLASDGIYNHAIQAODGS
FTSDIMTVAYVRD"

/complement(3016 . 3285)
/gene="YP03219"
/complement(3016 . 3285)
/gene="YP03219"
/note="no significant database matches"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="CAC92454.1"
/db_xref="gi:15981155"
/db_xref="SPTREMBL:Q8ZC11"
/translation="MEGSVMGISSEESITRTLTNEKNAGVSAKVAAIIISAVYTILMFE
YKHDELGYTLTAGDFLVLSFTIPWMKKRQRVKSYKNOLOOIEEDKTV"
/complement(3088 . 3138)
/gene="YP03219"
/note="one of 2 probable transmembrane helices predicted
for YPO3219 by TMHMM2.0"
/complement(3151 . 3216)
/gene="YP03219"
/note="one of 2 probable transmembrane helices predicted
for YPO3219 by TMHMM2.0"
complement(3468 . 4226)
/gene="YP03220"
complement(3468 . 4226)
/gene="YP03220"
/note="similar to Escherichia coli hypothetical protein
YCGR_SW_YCCR_ECOLI (P76010) (244 aa) fasta scores: E(): 0,
40.7% id in 241 aa"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAC92455.1"
/db_xref="gi:15981155"
/db_xref="SPTREMBL:Q8ZC10"
/translation="MGPVSETSKQPVPKNNKAICAILRIDIQNKIDTAVMWTXRGFFIT
SRILDIOETNQPIFDPSVENENVALGAEOITIIVEPTGAKLFCTCNKLKHVEYLTS
LPASFASIPDOYFIOREREYRVSI.FPMFLAYCGSGPDGTGYRTLADISIGMSGYLYE
AMKGSEFPLOGSVLRDAADVLDGGGLPKDLGITRALDKQVVANNKMELTYQLSPEK
FLRLSPIDKEKLGRIAFIFELEKOOTRAKRKPBEGL"
/complement(4554 . 5813)
/gene="YP03221"
/note="proca"
complement(4554 . 5813)

CDS

/EC_number="1.2.1.41"
/note="Similar to Serratia marcescens gamma-glutamyl
phosphate reductase ProA SW_PROA SERNA (P17857) (417 aa)
fasta scores: E(): 0, 85.1% id in 417 aa, and to
Escherichia coli gamma-glutamyl phosphate reductase ProA
SW_PROA ECOLI (P07004) (417 aa) fasta scores: E(): 0,
72.4% id in 417 aa"
/codon_start=1
/transl_table=1
/product="gamma-glutamyl phosphate reductase"
/protein_id="CAC92456.1"
/db_xref="gi:15981157"
/db_xref="SPTREMBL:Q8ZC09"
/translation="WNILEHWGAKAQSQWLAMLSTAKKNQALAVIANLLSESQTTI
LOANRDMAARBSGMSEALDRILLTPPARLAATANDVRCRLNDPEGVYIDSOLDVD
SGIKERRVERPVGVIYEARNPNVITDVASCCLKTNNAVTRGKETHTYNQNVQAWVN
IORALEGCOPPAAVAOATESPDROLNETLRDEYVMULPRGASLHKACRESTPIB
VITGTCIGHCFUDENDADEFKALIYENAKQRSANSLETILPHOVAKTPFILPAS
ARMARFGYTHLASPLANPYLDADGKAKVANAZAAYDDEMSJLDAVINITYIDIADAITH
IREHETSJDSDLITRSISHAYEVPAWDVASVYVNASRTFDGGQFELGABVAVASTOK
LARHPQMCLIDLITTKYMKYGDDDIURS"
/complement(4572 . 5744)
/gene="YP03221"
/note="pfam match to entry PF00171 aldehyd, Aldehyde dehydrogenase family, score -217.90, E-value 0.00013"
complement(4776 . 4841)

misc_feature

misc_feature

misc_feature

-35_signal
 1338..1343
 /note="potential site for katf gene"
 -10_signal
 1359..1364
 /note="potential site for katf gene"
 gene
 1399..1433
 /gene="katf"
 1399..1433
 /gene="katf"
 /codon_start=1
 /transl_table=11
 /protein_id="AA17876.1"
 /db_xref="GI:476826"
 /translation="MSQNTLKVHDL"
 BASE COUNT 381 a 365 c 369 g 318 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.000263 Length: 1433
 Score: 213.50 Matches: 87
 Percent Similarity: 42.01% Conservative: 55
 Best Local Similarity: 25.74% Mismatches: 137
 Query Match: 12.78% Indels: 60
 DB: 1 Gaps: 12

US-10-018-706-2 (1-322) x ECOPEMNLDP (1-1433)

QY 21 ILEPHEGLYVALLIETHIRHCYAILLEUVALAGLYCYALASERYLSPROTHRYASN 40
 DB 382 GTCTACACCGCAGCAGCAATTCAGCCGCTGACAGACCAAAATTCAGCGTACTCA 441
 QY 41 SERHSEGLYSERGLYSERHIAHYTHSERGLYSERGLYLEU-NAILEGLYSE 60
 DB 442 ACAACCCGCAATTCAGCCAGTGCAGCCAGTGCAGCCGCTGACAGTGCAGTGAACG 501
 QY 60 RGLNALILETHRASPSEGIN-----GLYVALPRO-----ASNAR 72
 DB 502 ACGCATGCTGTATACCGTCACTATGGGAACATTCGGAAGCAGTTATACGCGCAGTAC 561
 QY 72 GTYCTGVALIYGLINGLYAPRTHVALSERIYLEALAGINATGYTGYLEUASNT 92
 DB 562 CTATACCGTGAATAAAGCGCACACACTTCTATATGCGCTGATTAAGTCAACGATTT 621
 QY 92 PARGUILLEGLYHISILEASNALEUASERSETYRTHRIETRYTHIRGLYGINTR 112
 DB 622 CCGTACCGCTGCTCAGCGCAACATATTCAGGCACCAATACCGCTGAACTGTGCTCAG-- 679
 QY 112 PLEUTHIRLEUTRPSERGLYASP-LEUTYVALATRGJLUARSERILSESERGLYALA 132
 DB 680 -----ACCTTGACAGGTGGTATATGCTTCCGGTACCCCA-----T 714
 QY 132 SATHRALAHETHRPSERPROVALAVALAGINSESRARYPROPROVALAGINLN 152
 DB 715 CACTGGCGGAAATGCCATTACCC-----AGCCCGACGAGCAGACGA 756
 QY 152 TSPROALAVAGINLYPRO-----THRPROVALAVALAVALYLYSPROTHR- 169
 DB 757 AGGA-GTTGTGATCAACGCTGCACAAATTCACCGTTGCTGTGCTGCACACCGACA 815
 QY 170 -----P 170
 DB 816 TTACGTATTCGAGTCTTCGGGTGACAGAGTGTACAAATGTGCGCAACAACAGC 875
 QY 170 ROTHPROPROVALAVALAGINLPROVALAVALAVALAVALAVALAVALAVALAVALA 190
 DB 876 CAATCTGACACCAACCGTCAACGCTGTAAACGATCAACAGCAACGACGCGGA 935
 QY 190 HEALATHNGLY-----SERSERGLYVALMETGLINPHEATGYRPROVALAGLYLAT 207
 DB 936 CTGTGACAGTACATCAACAGTACGCTATCTCCACTGCGCTGCGC-----ACTG 989

QY 207 hrAenProValaValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSera 227
 DB 990 AGGCGCAAGATGATGAAACCTTTGGC---GCTTCTGAGGGGGC-----AAACA 1034
 QY 227 enGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrV 247
 DB 1035 AGGGGATGTATATGATGACGAGCAGCAAGACAGGCAATATGCGACAGATGCCCGC 1094
 QY 247 alIleGlnAlaAspHisAsmMerAspGlyAlaSer-----IleValIleGlnHisThrA 265
 DB 1095 TTGTTATGCTGTGATGACCGCTGCGCGCTTACCGTATTCGATATTCATCAACCAATATG 1154
 QY 265 enGlyPheValSerSerTyrIleHisIleYsaAspAlaGlnValIleThrGlyAspThrV 285
 DB 1155 ATGATTACCTGAGAGCGCTGACGCCATTAACGACACAAGTCGTGTCGGGGAACAACAAGAG 1214
 QY 285 alArgThrGlyGlnArgIleAlaSerMetIlysaAsnGlnProSerGlyAlaAlaLeuPheG 305
 DB 1215 TTAAAGCGGGGCAAAAATAGCCGACATGCTGACACCGGAACCAAGTTCAACAGCGCTTGC 1274
 QY 305 luPheArgIleSerArgAsnGlyValIleValAspProLeuThrValIleu 321
 DB 1275 ATTTGAAATTCGTTACAGGGGAATCCGTAACCGCTGCTGTTATTTG 1324
 RESULT 29
 AE005502/c
 LOCUS
 DEFINITION
 Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 121
 OF 290.
 ACCESSION
 AE005502 AE005174
 VERSION
 AE005502.1 GI:12517193
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli O157:H7 EDL933.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE
 1 (bases 1 to 10573)
 Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.V., Davis,N.W., Lim,A., Dimlantia,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)
 JOURNAL
 MEDLINE
 PUBMED
 21074935
 11206551
 2 (bases 1 to 10573)
 Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.V., Davis,N.W., Lim,A., Dimlantia,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 JOURNAL
 FEATURES
 source
 1..10573
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohemorrhagic"
 misc_feature
 <1..601
 /note="O-island #110; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
 88..528
 /gene="Z4048"
 88..528
 /gene="Z4048"
 /note="putative regulator; Not classified"
 /function="Residues 12 to 146 of 146 are 100.00 pct identical"

to residues 1 to 135 of 135 from Genepet 118 :
g14887558[emb|CAB43501.1] (A0006210) transcriptional
regulator protein [Escherichia coli]
/codon start=1
/transl_table=11
/product="putative regulator"
/protein_id="AAG57849.1"
/db_xref="gi:12517194"
/translation="MIAADSRGVMTMLPKAKFHOQLFOGHAAWGHLPDITQPO
VAWAKLADKSGIEGVLEATEKAVSTKATILAKMKNELTAREHDADKRRRFVWL
TAEGRKLAIAPIFGDSYDEERLGRLSAEQELFVQLVRKMMNT"
complement (648. .1640)
/gene="rpos"
/note="24049"
complement (648. .1640)
/gene="rpos"
/note="24049"
complement (648. .1640)
/function="enzyme; Global regulatory functions"
/note="Residues 1 to 330 of 330 are 99.09 pct identical to
residues 1 to 330 of 330 from Escherichia coli K-12 strain
MG1655: B2741"
/codon start=1
/transl_table=11
/product="RNA polymerase, sigma S (sigma38) factor;
synthesis of many growth phase related proteins"
/protein_id="AAG57848.1"
/db_xref="gi:12517195"
/translation="MSQNTLKVHDLNEDAEFDNDNRVAFPEKALVEEPESDNLAEE
LLSQATQVRNADATOLYGEIGYSLTAEEVYFARALAGVASRRRIEENILVY
VLIARQKNGVLDLIEEGNGLIRAVEFDEGRGPRSTVATWIRQTIERRALN
OITRTLRPIHYKELNVLYLTARESLHDEPSELAEDQLPVDVDSKRMFLNER
ITSVTPRIGYSEKALDILIDAEKENGPEDTQDDMKQSVKMLPELNKQEVLAR
RGLGLVAAITLDEGRIEIGLTRERVQIQVEGRRLREILQGNIEALFRE"
complement (1703. .2842)
/gene="rnpdp"
/note="24050"
complement (1703. .2842)
/gene="rnpdp"
/note="24050"
complement (1703. .2842)
/function="membrane; Macromolecule synthesis,
modification: Lipoprotein"
/note="Residues 1 to 379 of 379 are 99.73 pct identical to
residues 1 to 379 of 379 from Escherichia coli K-12 strain
MG1655: B2742"
/codon start=1
/transl_table=11
/product="lipoprotein"
/protein_id="AAG57849.1"
/db_xref="gi:12517196"
/translation="MSAGSPKTVYRIALSLVSLMTLAGSDTSNPAPVSYGNAP
AHTNSGALITPPKMGITSTAYQPOIIPVQPOQIQAQPOIIPVQVAAQPOWENG
RIVNRQYGNITPKSGYSGSTAYVKGKDLFYIAITQAGDEQALQNNILAPALVNG
OTLQVNGASGRTITGNAITQADAAQOCGVIRKPNQSNVAVASQPTIYESSGEQA
NMLPNKRTATTVTAPVTPVATTEPEVSTSTSPSTIATWAPTEGKYLETFGASE
GNGKIDIGAKGALIIATDQRRVYVAGNALRGYGLIITKNDVYLSAVAHNDMLV
RDOQEVKAGOKIATGVTGSTRHFEIRFKGKSVNPLRLTPQR"
complement (2982. .3608)
/gene="pcm"
/note="24051"
complement (2982. .3608)
/gene="pcm"
/note="24051"
complement (2982. .3608)
/function="enzyme; Macromolecule synthesis, modification:
Proteins - translation and modification"
/note="Residues 1 to 208 of 208 are 100.00 pct identical
to residues 1 to 208 of 208 from Escherichia coli K-12
Strain MG1655: B2743"
/codon start=1
/transl_table=11
/product="L-isoaspartate protein carboxylmethyltransferase
type II"
/protein_id="AAG57850.1"
/db_xref="gi:12517197"
/translation="WVSRRVVALDQLRAQGIODEQVINALAAVPRKFEVDEAFQKA
WNINLPIQGGITQSPVWAMVTEILBLTQSRVLEIGTSGVOTAILAHLVQVDS
VRIKIQIQGARRRKNTLDLHNVSRTSHDQNGQVQARAPPAIIIVTAAPPEIPALMT

gene
CDS
gene
CDS
gene
CDS
gene

QJLDBGILIVPEGEHNYLKRVRNRGGEEFIIDVEAVRFVPLVKGEIJA"
complement(3602. .4363)
/gene="sure"
/note="24052"
complement(3602. .4363)
/gene="sure"
/function="phenotype; Not classified"
/note="Residues 1 to 253 of 253 are 100.00 pct identical
to residues 1 to 253 of 253 from Escherichia coli K-12
Strain MG1655; B2744"
/codon_start=1
/transl_table=11
/product="survival protein"
/protein_id="AAC57851.1"
/db_xref="GI:12517199"
/translation="WRILISNDGCVNAPGIOTLAIRERADQVYVAPPNREGASNS
LITSSSLATFEPFGDIAVWGFTPTDCVYLGVNALRRPFDIYSGINAKPNIGDYVI
YSGVYAAMBRRLGFALPAVSLEDHGKRIIPAAVTSTIIRALCKEPIRGTIRLINTV
PDLPDIQIKGRIVTRICRHPRADVITQQDPKRGITLLWIFPGKCDAGETIPAAND
EGYSTIPLHDAVLPAHSADQVSDMLNSVGVTGW"
complement(4344. .5393)
/gene="ygdO"
/note="24053"
complement(4344. .5393)
/gene="ygdO"
/function="putative enzyme; Not classified"
/note="Residues 1 to 349 of 349 are 99.42 pct identical to
residue 1 to 349 from Escherichia coli K-12 strain
MG1655; B2745"
/codon_start=1
/transl_table=11
/product="putative hydrogenase subunit"
/protein_id="AAG57852.1"
/db_xref="GI:12517199"
/translation="MIEFDNLTYLHGKPOGGGLKANPEDPVVVVEDLGFEPPDGCEHI
LVRLIKGCNCNFRFAVDALAKELKHAREVSPAGOKDGAVTBOMLCARVSGKEMPDL
AFOLGCGCNFRFAAHKKRKIRGALKGNAPFLVREVNRDDEORLDICVKGPYNF
GVNDPDIQIALSRGSGFVAITTEIIVEIORVNDKIMTALLPSGSEGTORBALAFQC
AAVAAEITQLLIIVREYEAARRRMILYPQLSNMMWDVTVEIFLPLPAGSFRTSYV
RELNIITDDIAHAIR"
complement(5390. .5869)
/gene="ygbB"
/note="24054"
complement(5390. .5869)
/gene="ygbB"
/function="orf; Unknown function"
/note="Residues 1 to 159 of 159 are 100.00 pct identical
to residues 1 to 159 of 159 from Escherichia coli K-12
strain MG1655; B2746"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAG57853.1"
/db_xref="GI:12517200"
/translation="WRIGHGFDVAHFGEGEIIIGGYRIPEKGLIASHDGDVALHAL
TDALGAALADIGLKFGDTPDKADSRLLEARNRIQAQKYTGANDVITIAOAQ
PKMLPHIRQMWLFVAIEDGICMDDVNVKAITTTXLTGFTGSGEIACBAVALIIVATK"
complement(5869. .6579)
/gene="ygbB"
/note="24055"
complement(5869. .6579)
/gene="ygbB"
/function="orf; Unknown function"
/note="Residues 1 to 236 of 236 are 99.15 pct identical to
residue 1 to 236 of 236 from Escherichia coli K-12 Strain
MG1655; B2747"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAG57854.1"
/db_xref="GI:12517201"
/translation="MATTHADVCVAPAAGGRRMQTECPKQYIISIGNQTILHESVTA"

sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

SOURCE

Location/Qualifiers

1..11457

/organism="Escherichia coli K12"

/strain="K12"

/sub_strain="WG1655"

/db_xref="taxon:83333"

75..1439

/gene="b2740"

75..1439

/gene="b2740"

/function="putative transport; Not classified"

/note="0454; This 454 aa ORF is 34 pct identical (9 gaps) to 444 residues of an approx. 456 aa protein GMP_BACU1

SW: P46832"

/codon_start=1

/transl_table=1

/product="putative transport protein"

/protein_id="AAC75782.1"

/db_xref="GI:1789097"

/translation="MSSTTLCIALAGVIMLLIVIKAKQPFVALLVSLVLAAGIPAGEVGKVMAGMGVIGSVTIIIGAMGRMIHSGGSELANYSRRLGPKRTI

ALTLAEPFIPVPDPVIGIILAIPIYGPVAKKISPLKGLPAVIGMLVHAAPV

HGPVAAAGILHADIHMTIIGIISIPGVGVPAKTIKKROYAMSVELBEMOLA

PASREGATKSDKINPVALSVLSIVIPALIMAGTYSATIMPSHPLGGLTILIS

PMVIMTAVIAPFLIALRGMVLOHTSDINGSLPTPAVILVPGAGGVGKVLVS

GVGLANMLQMDILPLRALPFIISILPASGSATVALITGGILSEAVNGLPIC

VLTLACFGSGSHINDSGFIYTKILGSLVADGLKTMVLTITLIGFPGILTWCY

WAVI"

/complement(1533..2525)

/gene="rpos"

/note="b2741"

/complement(1533..2525)

/gene="rpos"

/function="enzyme; Global regulatory functions"

/note="F330; 99 pct identical to RPOS_ECOLI SW: P13445; CG

Site No. 18208; alternate gene names katF, appR, sigS,

otax, nur"

/codon_start=1

/transl_table=1

/product="RNA polymerase, sigma S (sigma38) factor;

synthesis of many growth phase related proteins"

/protein_id="AAC75783.1"

/db_xref="GI:1789098"

/translation="MSQWTLKVDHEDAEFDEGVEVDEKALVEQEPNDIAEEEL

LLSGATRVADATOLYGEIGVSPILTAEEVEYFARALGDVARSRRMESNRLV

VKIRRYGNRGALDLDEGNLGIIRAVEKDFPGRFSTVATWVRIOTIERALIN

OTRIPLRPIHIVKELNYLRTARELSHLDHESAEIEAQLDQVDSKMLNER

ITSDPLGDSSEKALDILDAKENGPEDTODDMKOSIVKMLPELNAKORELAR

RFGILGYAATLEVGREIGLTRERVRIOYEGRLREILITQGLNEALFIRE

/complement(2572..2603)

/note="factor Sigma70; predicted +1 start at 2865614"

/complement(2588..3727)

/gene="hlpD"

/note="b2742"

/complement(2588..3727)

/gene="hlpD"

/function="membrane; Macromolecule synthesis,

modification: Lipoprotein"

/note="F379; 100 pct identical to NLPD_ECOLI SW: P33648;

CG Site No. 33225; ORFX of D17549, has different start due

to frameshift in D17549"

/codon_start=1

/transl_table=1

/product="lipoprotein"

/protein_id="AAC75784.1"

gene

CDS

/db_xref="GI:1789099"

/translation="MSAGSPKFTVRILALSLVSLWAGCSDTSPNPAPVSVNGAP

AMNSGMLITPPKMGKTSTPAQOIPVOVQOIQOPOIPVOVQOIPVOVQOIPV

RIYNNQYNGIPKSGSYSTYKADDTLEYIMTIGNDPRDLAONNIIQAPYALNG

OTLOVANGSTPIITGNALTOADAAGVVIKPAONSVYVAGOPTITYESGEOGA

NKLLPNKPPATTVTAIVYPTSTTEPTVSTSTSTSTSTSTSTSTSTSTSTST

CGKJDIAGSKKQALITADGPRVAGNABRGYMLIIKHNDYLSAYAHNDTMLV

RQOEVYKAGCKITMGSTGSTRHFEIRYKKSVPNPLRYLPGR"

/complement(3867..4493)

/gene="rpm"

/note="b2743"

/complement(3867..4493)

/gene="rpm"

/EC_number="2.1.1.77"

/function="enzyme; Proteins - translation and

modification: 100 pct identical to PMT_ECOLI SW: P24206

but includes initiator met; CG Site No. 33221"

/codon_start=1

/transl_table=1

/product="L-isoaspartate protein carboxylmethyltransferase

type II"

/protein_id="AAC75785.1"

/db_xref="GI:1789100"

/translation="MVSRRVQALIDQLRAQIODEQVNLAAVPREKVEAEQKA

MDNIALPIGQGTISQVYVAVMTLELTPQSVLEIGTSGVQTAIHLVQVHS

VERIKGIQOGARRRLKDLHNSTHGDQMGQVAPDAIVTAAPPEIPTALMT

QLDEGGILVPGVEHQYLKRVRRRGGEFIIDTVAVRFPVLKGEIA"

/complement(4487..5248)

/gene="sure"

/note="b2744"

/complement(4487..5248)

/gene="sure"

/function="phenotype; Not classified"

/note="F253; 100 pct identical to PIR: I69732; alternate

name ybcC; ORF of L07942, has different start due to

frameshift in L07942"

/codon_start=1

/transl_table=1

/product="survival protein"

/protein_id="AAC75786.1"

/db_xref="GI:1789101"

/translation="MRLISNDGCHAPGIOTAKAREADVQVYVAPDRREGASNS

LTSSSLRTTFENGDIADVOMGTPTDCVYGVNAKRPDIYVSGINAGPNDGYI

YSGTVAAMEGRH/GFPALAVSLDGHCHYDTAAVTCSTIALCKEPLRTGTLINIV

PDLPLDIKIGIRVRCGRHPADQVLPQDPKRNLTWISGPGKCDAGGTDPAAVD

EGYVSTIRPLHVDLTAHSAQDVSDMNSVGVTQW"

/complement(4535..4563)

/gene="sure"

/note="factor Sigma70; predicted +1 start at 2867577"

/complement(5229..6278)

/gene="ygbO"

/note="b2745"

/complement(5229..6278)

/gene="ygbO"

/function="putative enzyme; Not classified"

/note="F349; was f292 and f255 before splice; This 349 aa

ORF is 44 pct identical (8 gaps) to 338 residues of an

approx. 344 aa protein y701_HABIN SW: P44039"

/codon_start=1

/transl_table=1

/product="putative hydrogenase subunit"

/protein_id="AAC75787.1"

/db_xref="GI:1789102"

/translation="MIFEDULTYHKGPGSTGLKANKPDPVVEDIGRPDGGHET

LVRILKGCNTRFADALAKFLKHNKAEVFAQKDHATBQDLCKRYGKMPDHS

AFQEGGVAVYAHKRLKRLGAKNAFTLVAEVNRDVEDRGLDLYKVCVNPV

GAQRFIGISNLOGAQNTNTPVRNRKRSFMSAKALRQIVAEKRAKDUNO

VVDSDALQLAGRSWFAVTEELAEIQRVNDKMLMTTALPSSGEGTORALAFEG

Alignment Scores:
Pred. No.:
Score:

0.0026 Length: 11457
213.50 Matches: 87

Percent Similarity: 42.01%
 Best Local Similarity: 25.74%
 Query Match: 12.78%
 DB: 1
 Gaps: 12

US-10-018-706-2 (1-322) x AE000358 (1-11457)

QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLeuProThrValAsn 40
 Db 3542 GTCTACAGCGAGCAACCAATTACAGCGGTGACAGACACCAAAATTCAGGCTACTCA 3483
 QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyLeu-AlaIleGlyse 60
 Db 3482 ACAACCGCAAAATCCAGCGACGACAGCTAGCTCAGACGCGGTACAGATGAAACGCG 3423
 QY 60 rGlnValIleThrAspSerGln-----GlyValPro-----AsnArg 72
 Db 3422 AGCGATCGCTATTAATCCGTCACTATGGAAACATTCGAAAGCACTTATACGCGCAGTAC 3363
 QY 72 GTTCGlnValIleGlySerGlnGlyAspThrValSerIleLeuAlaGlnArgIleuAsnTr 92
 Db 3362 CTATACCGTGAATAAAGCGACACACTTTCATATCCGCTGATCTACGCAACGATTT 3303
 QY 92 PArgGluIleGlyHisIleAsnAsnLeuAsnSerSerIleThrIleIleThrGlyGlnTr 112
 Db 3302 CCGTACACCTTCCTCAGCGCAACAATATTCAGGACCATACGCGGTGAACGTTGTCAG-- 3245
 QY 112 PleuThrIleuThrSerGlyAsp-LeuIleValArgGluArgSerIleSerSerGlyValA 132
 Db 3244 ---ACCTTGACAGGTGGGTAAATGCTCCGTAACGCCA-----T 3210
 QY 132 snThrAlaHisIleThrProSerProValAlaValGlnSerSerArgProProValGlnGln 152
 Db 3209 CACTGGCGGAATATGCATTATCC-----AGCGCAACGACGACGACCA 3168
 QY 152 IspProAlaValGlnIlePro-----ThrProProValValValValIleValysProThr- 169
 Db 3167 AGCA-GTTGTGATCAAGCGCTCACAAATATCCACGTTGCTGTCGTGCGAACGCA 3109
 QY 170 -----P 170
 Db 3108 TTACGATATTCGAGCTTCGGGTGAACAGAGTGTACAAATGTTGGCGCAACAAAGC 3049
 QY 170 rOThrProProValValGlnGlnIleProAlaProValAlaProProValThrGluAlaProp 190
 Db 3048 CAATCGCAACCAAGGTCAACGCGCTGTAAAGGTACCAACGACGACGACGACGCGCA 2989
 QY 190 heaIaThrGly-----SerSerGlyValMetGlnPheArgIleProValGlyAlaI 207
 Db 2988 CTGTACAGAGTACATCAACACAGTACCTATCTCACTGCGGCGGTGCGCG-----ACTG 2935
 QY 207 hrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerA 227
 Db 2934 AGGCGCAAGTATCAAACTTTGGC---GCTTCGAGGGGGC-----AACG 2890
 QY 227 snGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrV 247
 Db 2889 AGGGATGATATTCGACGACGCAAGACGCAATATTCGACCGCGACGACGCGCGCG 2830
 QY 247 aIleGlnAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisI 265
 Db 2829 TTGTTTATGCTGTAAAGCGCGGTGCGCGCTACCGTAATCTGATTCATCAACATTA 2770
 QY 265 snGlyPheValSerSerIleHisIleValAspAlaGlnValIleThrGlyAspThrV 285
 Db 2769 ATGATTACTCTAGTCCCTACCGCATTAACGACACATCTGCTGCGGAACAACAAG 2710
 QY 285 aIArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyValAlaLeuPheG 305
 Db 2709 TTAAGCGCGGCAAAATTAAGGACCATGCGGAGACCGGAACGTTCAACAGCTTGC 2650
 QY 305 IupPheArgIleSerArgAsnGlyValIleValAspProLeuThrValIleu 321

Db 2649 ATTTGAATTCGTTACAGGGGAATCCGTAACCCGCTGTTATTG 2600
 RESULT 31
 ECU29579/c
 LOCUS ECU29579 72221 bp DNA linear BCT 01-JUL-1995
 DEFINITION Escherichia coli K-12 genome; approximately 61 to 62 minutes.
 VERSION U29579
 KEYWORDS U29579.1 GI:882596
 SOURCE
 ORGANISM Escherichia coli.
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE
 AUTHORS 1 (bases 1 to 72221)
 TITLE Plunkett, G.P.
 JOURNAL Direct Submission
 Submitted (20-JUN-1995) Guy P. Plunkett, University of Wisconsin,
 Laboratory of Genetics, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined as part of the E. coli Genome Project
 (Frederick R. Blattner, director) at the University of
 Wisconsin-Madison. Supported by award HG00301 from the NIH Human
 Genome Project. The entire sequence was independently determined
 from E. coli MG1655; overlaps with other sequence determinations
 are annotated. This entry should be considered somewhat
 provisional; it will be updated and merged with others at a later
 date.

FEATURES
 source
 location/Qualifiers
 1..72221
 /organism="Escherichia coli"
 /strain="K-12"
 /sub_strain="MG1655"
 /db_xref="taxon:562"
 /map="approximately 61 to 62 minutes"
 /note="this sequence comprises the following lambda
 clones: DB609 (EC30K46-3), DB612 (EC19-195), DB615 (EC27-
 240), DB618 (EC22-172), and DB620 (EC18-1463); MJ3Janus
 vector was used for subcloning"
 1..2371
 /note="similar to GenBank Accession Number J02708
 (ECOGUT)"
 39..818
 /gene="gucD"
 39..818
 /gene="gucD"
 /EC_number="1.1.1.140"
 /note="CG Site No. 153; alternate gene name srlD"
 /codon_start=1
 /transl_table=1
 /product="glucitol-6-phosphate dehydrogenase (sorbitol-6-
 phosphate 2-dehydrogenase)"
 /protein_id="AA6924.1"
 /db_xref="GI:882597"
 /translation="MNQVAVVTGGGQTGAFLCHGLAAEGYVAVVDIQDKAANVAQ
 EINAYESMAVYAGGADATSESVTALSRGDEIFGRVDLVYAGIAKAAFISDFOL
 GDFPRLQVNLVGYFLCARERSRLIMDIGRIQLINSKSGKSGKNSGSAKFG
 GVGLTGLADLAIEYGTIVSIMGNLTKSPQSLIPQVATKLGIRKDYQEQYVYIDK
 VPIKRGCDYQDVANMLFYASPKASYCTGGSINTGCGVM"
 572..2371
 /note="similar to GenBank Accession Number X13463
 (ECGUTM) 1..1800"
 924..1283
 /gene="gucM"
 924..1283
 /gene="gucM"
 /gene="gucM"
 /function="glucitol operon activator"
 /note="CG Site No. 33198; alternate gene name srlM"
 /codon_start=1
 /transl_table=1
 /protein_id="AA6924.1"
 /db_xref="GI:882598"
 /translation="WVSALITVAIVANCAOLALGKQWISRPNRAFDTLCCGGRVGVG
 SSGRFPKRVVVAALDDQORIVDTLFMKGLGVFARPKIPAITGMHAGDLOPDVIFPH
 DPLSQNALSLAKLKRG"

```
gene
1350..2123
/gene="gucR"
CDS
1350..2123
/gene="gucR"
/function="glucitol operon repressor"
/note="CG Site No. 152; alternate gene name srlR"
/codon_start=1
/transl_table=11
/protein_id="AA069216.1"
/db_xref="GI:882599"
/translation="MKPRQQAILEYLOKQCKSVBELAOFDTTGTTRKIVILE
HAGVIRTYGGVIVNKESDPIDHKTILINHKBELIAEAVSFTHDSIILDAGST
VLOWVPLSRFNNTIVMTNSLHIVNALSELNEQITLIPGTFPKKSKSPHQAENA
FEHPTFDLFWGIDGIDLAGVTFNEVYTVSKAMCAAREVILMADSKFRKSPNV
VCSLESVDKLTIDAGIDPAFRQALEEKIDIVITGESNE"
2121..3054
/note="similar to GenBank Accession Number X51361
(ECGUT0)"
misc_feature
2155..3081
/gene="gucO"
CDS
2155..3081
/gene="gucO"
/note="alternate gene name srlQ"
/codon_start=1
/transl_table=11
/protein_id="AA069217.1"
/db_xref="GI:882600"
/translation="MLRQEPASRLPERLGDPPVRAANILHCCEKVAVVSGIGKSGHTG
KRIATLSTCTPAPFVVPARALHGDLCMISRPVMLPTISGAKKEDILILPRLDG
SLIALAMGKPSPLGIAAKVLDISVERACPMILASTSVNTVMGDLAMVMO
ARGREBPASHPAGLIGARILNKHMERDAILROVALTASVMDMLSLRSGLS
LVAVDAAQQVQGVTTDDIRMLVCGALTPPVNEAMTVGGTTLQSSRAIDAKELL
MKRTIAFVVDENKLTGALIDPFGAII"
complement(3078..3590)
/note="alternate name ygaA"
/codon_start=1
/transl_table=11
/label="ORF_f170"
/protein_id="AA069218.1"
/db_xref="GI:882601"
/translation="MLAGRFADLFHRLSVPLSVPLREGRDDVILLAGYFCECRL
RQGLSRVVISAGRNLIQHYSPFGNRELEHRIHRAVILARTRSGDVIIIEAQHAF
BEVTLPTREVAAPVVKONLREATEAFQRETRQALQNHNMWAAACAMLETDVANLH
RLAKRLGLKD"
complement(3610..4668)
/note="alternate name ygaB; ORF_f352"
/codon_start=1
/transl_table=11
/protein_id="AA069219.1"
/db_xref="GI:882602"
/translation="MSSONDYQSGNRYNITYSQIDSEAMSPFVDVLANIAIELOGLI
GHODRFORLITLRQVLECDASALLRYSRQPIPLAIDGLADVGRPALIEGHRLE
ALRAGDVYRFPADSELPDPYGLI RQGESEVHACVGLPLPAGNLIAGLILDMOP
DOFVPSDEBELILALAAAGLSNMLIEOLSSONLGGDAPFAVAVOTOMIGLSPG
MTOLKREITVASDLNVLISGETRGEIKVAKIHEASPRVNPVLINCALESYV
AESRFGHVKGAFCGALISRSKGFPMANGTILFLDEIGELSLAQAKLRLVLYQYDIO
RQDDRCARVDRVYIGD"
4780..6219
/note="ORF_c479"
CDS
4780..6219
/codon_start=1
/transl_table=11
/protein_id="AA069220.1"
/db_xref="GI:882603"
/translation="MSIVKNNIHVWGQDWEVRDPFHCTEYTLKSSSYSLIREEK
TANAIDSLNGHHHEPMENNVKGTDTLIDIGKQOLIIVETMLMPSMWTYLTGCA
VLENDAPQGHYCDHLENDVDQTELEPCQCORYANILVTPSRLVTPKITEILLENL
PVDMLATSHGVARDNPTQIVELYKMAADYQEDRITTFYDMSNVTNRKIMADALIQGI
AETDPRVAVKIFNVASDKNEILITVFKSGVAVGISTNNANNMFKIAGLVEEMGLR
FANKKASAFSGMSGAGVADRSLSTRLODAGFMSLSLKAKNMPDDALALCEHREBI
AOMKALAPLQSTVNTVKEETSAITLADIGRMOCVCOMIYIDAKGEMODVAGT
PSEVVDNPLCECSIGKDVFEELASAK"
6216..7349
```

```
misc_feature
/note="ORF_c377"
/codon_start=1
/transl_table=11
/protein_id="AA069221.1"
/db_xref="GI:882604"
/translation="MNSGIVITGSGFPAQRLVKNIRKODATPILTLIAADSMDEVNKP
DLSHVISOGRADLTROTACGERAEQFNHILFPQVWTDIDAEARVYKSONMOYDK
LYIARGASCPVRPPGRELMLTNSQGEVACETOLRDAARVILVGGILGSELAMD
CPAGKVTITDNASTLASLMPPEVSRLOHRLITREMGVHLLSLOQLEKTSGIOA
TUDORNTIYDVAITANGAPETRLARASLTITNRGVCSYIOTSNTDLYALGDCAE
INCQVPELOPIOLSMVLAKNLIGNNTPKLPAMLYKTIPTPLPHLAEGETORQDLR
MONTREQSNVAVGVDADOLRAFYVSEDEMKAFGLKTLPM"
complement(6641..10718)
/note="similar to GenBank Accession Number D14422
(ECOHYA)"
misc_feature
6641..10718
/gene="hyda"
complement(7477..9729)
/note="hyda"
/function="transcriptional regulatory protein"
/note="CG Site No. 18232; alternate gene name hypF"
/codon_start=1
/transl_table=11
/protein_id="AA069222.1"
/db_xref="GI:882605"
/translation="MAKNTSCGVQLRIRGVQGVFPFVWQIAQQLNLHGVDVCDND
GVEVRLDEDETFVLOVHOCPPLARIDSVREPERFISQLPTEPTIROSQGMNTQI
VPDAATCPACIAEKNTGERRRYRPFINCHGCRFTIIRAMPEDRPFYMAAFPLCP
ACDEKVEDPDRRHAQPVACBEGCPLEWVSHGEHLEBAALQALIAQLMKGIYAI
KGIQGFPLCDARNSNAVATLRARKEPAPKPLVMLEVADGLPDAQQLTTPAPAVI
LVKKYVPELCDIAPLNEVGVLPANPQHLLOLQCPILVWTSGLSGKPRALIN
EOLADLOQIADGELINRDIYOMDSVYRSEGMRLRSRGVPPDALPFGKXNP
PVLCIGADLNKFLVGEQAVLSQHGDSDDIOQWMBEALRMQNIYDFEQYVY
HDNAPGVSSOMAREMMLPTQVLIHHHAHAAACLAHQPMNLALGCAAPATLSBGAALAE
NGLMWGECRLRVYRRECEHIGLPAVALPGGDIAAQKPMNLALGCAAPATLSBGAALAE
TASVQHWMSVLAIRERGINAPLASGCFPPVAALGCAAPATLSBGAALAE
LAASCHGVTRVTPRVNDOLATTPQOCLNMQCA PNOVARAFHDLAAGFALAME
QATIRGTTTLVFSQGVVHNLRLRLAHYLAADTLFLFPQSLPADGGLSLGCGVIAAA
RWLAGEVQNG"
complement(9882..10409)
/gene="hyda"
complement(9882..10409)
/note="hyda"
/function="electron transport from formate to hydrogen"
/note="CG Site No. 33521"
/codon_start=1
/transl_table=11
/product="4Fe-4S iron-sulfur protein"
/protein_id="AA069223.1"
/db_xref="GI:882606"
/translation="MNRFTIADASKICGCTCEVACVSHQENODCASLTPETELPR
HIVIGVNISTVYVRCEDAPCANVCENGAISRKGVHVAQECICQKTCVACPG
AMEVYVPEVINSAGLNVADRAKANKCDLCNREDDGPACMAACPHTALICVDNRKL
EOLAERKRRTALMF"
10433..14724
/note="similar to GenBank Accession Number M73326
(ECOASCBFG)"
complement(10558..11571)
gene
Alignment Scores:
Pred. No.: 0.0199 Length: 72221
Score: 213.50 Matches: 87
Percent Similarity: 22.07% Conservative: 55
Best Local Similarity: 25.74% Mismatches: 137
Query Match: 12.74% Indels: 60
DB: 1 Gaps: 12
US-10-018-706-2 (1-322) x ECU29579 (1-72221)
OY 21 IlephegiValIleThrThCysIleLeuAlaGlyCysAlaSerLysProThrTyran 40
DB 40873 GTCTACACGCGACGACCAATTCACCGGTGACGACGACCAATTCAGGCTACTCA 40814
```

Oy	41	serthserglyserglyserhiiaargthserglyserlyglyleu-Alallelyse	60
Dh	40813	ACAAACCGAATTCACCGCCAGCGGATGCTACGACCGGATACAGTAGGAAAAACGG	40755
Oy	60	rglnValIlethrAspserGln-----GlyValPro-----AsnAr	72
Dh	40753	ACGATGCGTCTATATACCGTCAGTATGGAGCAATTCCGAAGAGCAAGTTATAGCGCGAGTAC	40694
Oy	72	gTyrgInVallygInglyAspThrValSerlySillealagInAxyTlyglyLeuAsnTr	92
Dh	40693	CTATACCGGTAAAAAAGGCGACACACTTTCTATATCGCTGATTAATCGCAACGATTT	40634
Oy	92	pArgGlnIleGlyHisIleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTr	112
Dh	40633	CCGGAGCTTGCTCCAGCGCAACATATTCAGGACCAATACGCGTAAAGCTGGGTAG--	40576
Oy	112	pleuthrLeuTrpSerGlyAsp-LeuLyValArgGluArgSerIleSerSerGlyValA	132
Dh	40575	----ACCTTCAGAGTGGTATATGCTTCGGTACGCCAA-----T	40541
Oy	132	snThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGln	152
Dh	40540	CACGTGGGAAATGCGCATTAACC-----AGCGCGACGACAGACAGA	40499
Oy	152	isProAlaValGlnLysPro-----ThrProProValValValLysProThr-	169
Dh	40498	AGGA-GTTGTGATCCAAAGCTTCGCAAAATTCACACCGTTGCTGTGGCGTGCACACGACAA	40440
Oy	170	-----P	170
Dh	40439	TTACGTATTCGAGTCTTCGGGTGACAGAGTGTCTAACAAATGTTGCCAACAACAAC	40380
Oy	170	roThrProProValValGlnGlnProAlaProAlaProProValThrGluAlaProP	190
Dh	40379	CAACTGGACACAGGTACAGCGGCTGTAAAGGTATACAGCAACAGACAGACAGACCGCA	40320
Oy	190	healIatHrGly-----SerSerGlyValMetGlnPheArgTyProValGlyAlaT	207
Dh	40319	CTGTACCAAGTACATCAACCAAGTACGCGCTATCTCCACTGCGCTGGCCG-----ACTG	40266
Oy	207	hAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerA	227
Dh	40265	AGGCGAAAGTATGAAACCTTTGGC--GCTTGTGAGGGGGGC-----AACCA	40221
Oy	227	snGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrY	247
Dh	40220	AGCGGATTTGATATGCGACGACGACGAAGAAGACAGCAATTAATCGGACCGCAGATGGCGGG	40161
Oy	247	allieGlnAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrA	265
Dh	40160	TTGTTTATGTGTGTAAACCGCTGGCGGCTACAGGTATATGATTAATCATCAACAATAAG	40101
Oy	265	snGlyPheValSerSerTyThrIleHisIleLysAspAlaGlnValLysThrGlyAspThrY	285
Dh	40100	ATGATTTACCTGAGTGCCTACGCCCATACACACCAATGCTGTGTCGGGAAACAAGAAG	40041
Oy	285	alArgThrGlnGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheG	305
Dh	40040	TTAAGCGCGGCAAAAATAGCAGCACTGGGTAGCAACCGGACACAGTTCAACAGCGTTGC	39981
Oy	305	luPheArgIleSerArgAsnGlyValTyValAspProLeuThrValLeu	321
Dh	39980	ATTTTGAATTCCTTACAAAGGGAATATCCGTAAACCGCTGCGTTATTG	39931

RESULT 32

AP002562/c 270365 bp DNA linear BCT 07-MAR-2001

LOCUS Escherichia coli O157:H7 DNA, complete genome, section 13/20.

ACCESSION AP002562 GI:13362858

VERSION

KEYWORDS

Escherichia coli O157:H7 (strain:O157:H7, sub_strain:R1MD 0509952)

DNA.

ORGANISM	
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
Escherichia coli O157:H7	
REFERENCE	
AUTHORS	1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakiwa,C. and Shinagawa,H.	
TITLE	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak
JOURNAL	Genes Genet. Syst. 74 (5), 227-229 (1999)
MEDLINE	20198780
REFERENCE	2 (sites)
AUTHORS	Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.
TITLE	Comparative analysis of the whole set of RNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655
JOURNAL	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE	20557356
REFERENCE	3 (sites)
AUTHORS	Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakiwa,C. and Shinagawa,H.
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak
JOURNAL	Gene 258 (1-2), 127-139 (2000)
MEDLINE	20564182
REFERENCE	4 (sites)
AUTHORS	Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakiwa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shibata,T., Hattori,M. and Shinagawa,H.
TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL	DNA Res. 8 (1), 11-22 (2001)
MEDLINE	21156231
REFERENCE	5 (bases 1 to 270365)
AUTHORS	Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kengen-info.osaka-u.ac.jp/, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)
COMMENT	genome project.
FEATURES	
source	location/Qualifiers 1..270365 organism="Escherichia coli O157:H7" strain="O157:H7" /sub strain="RIMD 0509952" /db xref="taxon:83334" complement(435..1220) gene="EC83388" complement(433..1220) /gene="EC83388" note="similar to SSBB ECOLI g1 1788871 percent identity 98 in 261 aa (Conserved in E.coli K-12)" /codon_start=1 evidence=not_experimental transl_table=1 product="enhanced serine sensitivity" /protein_id="BAB36811.1" /db_xref="GI:13362859" translation="MTVMSEKVTLELDLKAAPRRHRAFPITILESTVWVGTAAGGEAVDSALDLQHWKEKDSTISVPFTTSDELAQQVVEBDAQVNVPRVLFEFNTIGTFLPANKLPGRKEFMPEIKSLILGEGNPNSSCEILLEGSLLSVAEPAPMIDSS
CDS	
gene	


```

Qy 92 parglurleaglhisileasnasleuasnsersertyrthriletyrthnglylntx 112
Db 214475 CCGGACCTTCTCAGCGCAACAATATTCAGGACCAATACCGGCGCAAGCTGCGCAG-- 214418
Qy 112 pleuthleutrpserglyasp-leulysvalarglnhargserileserglyvala 132
Db 214417 ----ACCTTACAGGCGGTAAATGCTTCGCGTACCGCA-----T 214383
Qy 132 anthralahisthrproserprovalalavalginsersearhproprovalaglnsln 152
Db 214382 CACTGCGCGCAATGCGCATTAACCC-----AGGCGGACGCGACAGACCA 214341
Qy 152 isproalavalaglnlyspPro-----ThrProProvalavalyslyspProthr 169
Db 214340 AGGA-GTTGTATCAAGCCTGCACAAATTCACCGCTTGCTTCGCGACCGACGACA 214282
Qy 170 -----P 170
Db 214281 TTACGTATTCTGAGTCTTCGGGTGACAGAGTGTACAAATGTGCCGAAACAACAGC 214222
Qy 170 rothrrproprovalavalaglnhproalaprovalalaprovalthrghulaprop 190
Db 214221 CAACGCGCACCGGTACAGCGCTGTACCGGTACCAACAGCAACCGACCGCA 214162
Qy 190 healtargly-----serseglyvalmetglnphargtyrprovalaglylat 207
Db 214161 CTGTACAGACATCATCAACGATACGCTATCTCCACCTGCGCGCGCGC-----ACTG 214108
Qy 207 hransprovalavalargharghglythralthrvalalaglyserthrvalthrsera 227
Db 214107 AGGCGCAAGGTATGAAACCTTTGGC---GCTTTCAGAGGGGGC-----AACCA 214063
Qy 227 smglymettrpserglyargaspglyaspheulleasnalaseranaglythr 247
Db 214062 AGGGGATGATATGCGGACGAGCAAGGACGCAATATTCGCGACCGCAGATGCGCGTG 214003
Qy 247 alileglnalaasphiasmetasrphglyalaser-----llevalileglnhistr 265
Db 214002 TTGTTATGCGCGGTACCGCGCTGCGCGCTACCGTATCTGATATCATCAACATATATG 213943
Qy 265 smglyphevalaserseertyrlehisilelysaspalaglnvalythrghlyapthrv 285
Db 213942 ATGATTACGTAGTGCTTACCGCCATACGACACAAATCTGCTCGGGAACAACAAG 213883
Qy 285 alaargthrglyglarglalelasermetylasnglnproserglyalalaleupieg 305
Db 213882 TGAAGCGCGGCAAAATAATACCAATGCGGTACGACCGGAACCGTTCAACGCTTGC 213823
Qy 305 lupheargileserargasnlyvaltyrvalaspProleuthrValleu 321
Db 213822 ATTTTGAAATTCGTTACAGGGGAAATCCGTAACCCGCTGCTATTG 213773

RESULT 33
AL627276 274050 bp DNA linear BCT 06-JUN-2002
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome, segment 12/20.
ACCESSION AL627276 AL513382
VERSION AL627276.1 GI:16503805
SOURCE Salmonella enterica subsp. enterica serovar Typhi.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE 1 (bases 1 to 274050)
AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,J., White,N.,
Farrar,J., Feltham,T., Hamlin,N., Haque,A., Hien,T.T., Holtzclot,S.,
Jagels,K., Krogan,A., Larsen,T.S., Leather,S., Mouton,S., O'Gea,P.,
Perry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,

```

```

TITLE Stevens,K., Whitehead,S. and Barrell,B.G.
JOURNAL Complete genome sequence of a multiple drug resistant Salmonella
MEDLINE enterica serovar Typhi CT18
PUBMED Nature 413 (6858), 848-852 (2001)
21534947
11677608
2 (bases 1 to 274050)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
Sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES
source
location/Qualifiers
1..274050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/strain="CT18"
/db_xref="taxon:90370"
197..201
/note="possible RBS"
208..897
/gene="STY2840"
/note="ung"
208..897
/gene="STY2840"
/EC number="3.2.2.-"
/note="orthologue of E. coli ung (UNG_ECOLI); Pasta hit to
UNG_ECOLI (228 aa), 91% identity in 228 aa overlap"
/codon_start=1
/transl_table=1
/product="uracil-DNA glycosylase"
/protein_id="CAD05831.1"
/db_xref="GI:16503806"
/db_xref="SPRMBL:O8XK6"
/translacion="MATELTHTDVLADBEKQPYINTLTATYAGRSQITTYPPQKDY
FNARFTBLGDAVKVILIGDPTFGPGAGHGLASVPGIAPPSLVNMYKELESLPG
FVRPARGVLESMVAGVLLNTVTVPAQGHASLSGWEFTDKVLSLINOHGEGV
FLWGSAAOKKGAIIIDPQRHILKAPAPSPASARHGFECNHFPALTQMLEQHEKTI
DWTPVLAPESE"
322..867
/gene="STY2840"
/note="STY2840"
/note="STY2840 match to entry PF00315 UNG, Uracil-DNA
glycosylase, score 451.20, E-value 9e-132"
376..405
/gene="STY2840"
/note="PS00130 Uracil-DNA glycosylase signature"
complement(1013..2050)
/gene="STY2841"
/note="YfiF"
complement(1013..2050)
/gene="STY2841"
/note="orthologue of E. coli YfiF (YfiF_ECOLI); Pasta hit
to YfiF_ECOLI (345 aa), 88% identity in 345 aa overlap"
/codon_start=1
/transl_table=1
/product="putative RNA methyltransferase"
/protein_id="CAD05832.1"
/db_xref="GI:16503807"
/db_xref="SPRMBL:O8XFT1"
/translacion="MNDELKNGSGKVAVVYRSDSDSKTHNRTGKGGKPAKSRT
DGGRRPARDERNNOSRDKHETSFWRTVSAAPGDEPEKVDHGSISSKSFIDPEVLR
QRAEIVYGCNACOLFQSRPDALIVAMVIOSTVPRFKALRWANRAHYVDA
ELKASGTEHNGGCVFLIKRNGTVAOWKQADOCVLALEBVANPHNGWNRSC
AHRGQALFSTTLPEKMYLVIGREVDYIPEARREDDLCVKINGTVESLWVSATGV
LMEWNRQKNA"
complement(1040..1462)
/gene="STY2841"
misc_feature
misc_feature
gene
CDS
misc_feature

```

```

/Note="Pfam match to entry PF00588 Spou_methylase, Spou
RNA methylase family, score 181.50, E-value 1.4e-50"
RBS
2244..2248
/Note="possible RBS"
gene
2254..2673
/Note="STY2842"
/Note="trixC; yfig"
/Note="STY2842"
/Note="STY2842"
/Note="Orthologue of E. coli yfig (THI2_ECOLI); Psta hit
to THI2_ECOLI (139 aa), 94% identity in 139 aa overlap"
CDS
/codon_start=1
/product="thioredoxin 2"
/protein_id="CAD05833.1"
/db_xref="GI:16503808"
/translation="MTVCTHCOAINRIPGRLQDAACGCGHELEFGEVINATGET
LDLKLKDLPLVDFWAPWCGPCNPAIFEDVAERSGKVRFKVNTAEARELSARF
GIRSPIMIRFKHGQVDMNGAVKAPFDSMLNEAL"
2353..2670
/Note="STY2842"
/Note="Pfam match to entry PF00085 thiored, Thioredoxin,
score 116.00, E-value 1.1e-33"
misc_feature
2419..2475
/Note="STY2842"
/Note="PS00194 Thioredoxin family active site"
gene
2746..3426
/Note="STY2843"
/Note="STY2843"
/Note="Orthologue of E. coli yfip ECOLI; Psta hit to
yfip_ECOLI (232 aa), 88% identity in 226 aa overlap"
CDS
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD05834.1"
/db_xref="GI:16503809"
/translation="MTNNAVLQRAERLARTPELARGNVRRCORCLLPKSCLCD
LTPSOAKSRCLVMEDEPMKPSNTGILADIDTAPAEKSRTEPOALLEVOHP
DYQPIVPAVAGAREVISTPAPGKPLFLMDGTAPAKMRKSPYDHDVVIS
VDLSRLSAVRLREIHAEOYCTAEVAIALLDLAGDPEAATISGEHPIRFKRYLAGKT
QHPGNVTA"
3480..5140
/Note="STY2844"
/Note="STY2844"
/Note="Similar to proteins from bacteria, archaea and
eukaryotes e.g. Entamoeba histolytica acetyl-CoA
synthetase TR:Q9NAT4 (EMBL:AF286346) (713 aa) Psta
scores: E(): 0.34,1% id in 721 aa and Giardia lamblia
acetyl-CoA synthetase TR:Q9YIN2 (EMBL:AF107206) (726 aa)
Psta scores: E(): 0.30,0% id in 700 aa
Orthologue of E. coli yfip ECOLI; Psta hit to yfip_ECOLI
(886 aa), 92% identity in 866 aa overlap"
CDS
/codon_start=1
/translation="MSQQLKALRPKSIAGVAGAKPRRAGTYLMMRNILAGFNGPV
LVTAPWAKVAVGMAPDASLPFTPDLLICTNARSRLALDLAGKGCCTITLSA
PTSOHELLASARHKMRLLGNSJGLAPWQGLNARSFVPIKQKLAFTISOAVS
NITLWAOQREWFYSYFIALGDSLDIDVELLDYARLDSKTAIILYELQSDARFV
SARSAKRNKPTLVISKRSPPAQRLLNTSAGMDPAWDAIGRAIGLRVODTHEPSA
VETLSHMRPLHGDRLMITSNGAPALALDELMSRNGATISSECTLOLRDTPHAI
DIANPLDCLDDASSEHYVKTLLILASODPALMVTGSSAALPGTESHALLETIKR
HRRGKFTLLTWCGEFSQEARLSEGLPTPRYRQPTITAFPMHVEYRNQQLR
EPALPSNLTNTVAEHLQEARLEAGATSLDTHETQPIIHAHYDHTPTWIASQAE
AIIAEQIGVAPALKRSPDIPKSEVQCVMLYLRTPASVQQAANAIPRVKAMFQA
RHHGLIVQSMNARAGAEIRVVVEHDPVPGPLIMGEGVEWRPEBOAVVALPPLMNM

```

```

LARYLVQGIKQKIRARSALRPDIIVGLSQLVQSNLIVDCEIQRDLIHPLASA
SEPTALDVTLLDIPFDGNSRLAVPYPHOLEVEMKNGRCLEPILPEDEPQIR
QPIAYOTKELLYRFPSEINEFHEDLANTQIDYDRENAFVVRMDNAEELIGVTR
AISDPNVAAEFVALVRSDLKGLGRIMKELIAYTRDHGLKRLNGITWPNRGNVA
LARKLGVQDIDQDEGIVGLTNLAKCDS"
5844..6071
/Note="STY2844"
/Note="Pfam match to entry PF00583 Acetyltransf,
Acetyltransferase (GNAT) family, score 59.00, E-value
2.9e-14"
gene
6255..7610
/Note="STY2845"
/Note="Psta"
/Note="Psta"
/Note="STY2845"
/Note="STY2845"
/EC_number="2.7.8.8"
/Note="Orthologue of E. coli psta (PSS_ECOLI); Psta hit
to PSS_ECOLI (451 aa), 94% identity in 451 aa overlap"
CDS
/codon_start=1
/product="CDP-diacylglycerol-serine
O-phosphatidyltransferase"
/protein_id="CAD05836.1"
/db_xref="GI:16503811"
/translation="MLSKFRNKHQOHLAQPKRISQSDVDDEFFYPATRETLLEKI
ASATORTICIVALEODPGGGLIDALYAAKRPBELDVAVDMHMAQGRIGAAAS
NTWADWYCRLAQENPGIDVPYGVPTNTRALGVLRKPGIIDDVSLVSGASNDVYL
HOHDKYRDRYQOLRNROMADIMFDWOTONLNMNGVNRFLDNTORPSKPIKNDIRLY
ROELRDASYHFOGDANDEOLSVTLVGLGSSLNKTIPELMCAEHLKICGPEFL
PAVLVNNIIOLEBDGKVEIIVGDKTNDYIEDEDFKIGALPYVEINLRFLEL
LOYVNTDQVLRVLRMKDDNTYHLKGMWVDKXMLLIGNNLMFRAMLDLENLILHD
PKQELARQKRELELIRHTTYIVKHYRLOSIDVADYVAKKLIRLRIRIDRLISRI
L"
6651..6731
/Note="STY2845"
/Note="STY2845"

Alignment Scores:
Pred. No.: 0.139 Length: 274050
Score: 210.00 Matches: 84
Percent Similarity: 38.67% Conservative: 56
Best Local Similarity: 23.20% Mismatches: 138
Query Match: 12.57% Gaps: 84
DB: 13

US-10-018-706-2 (1-322) x AL627276 (1-274050)
QY 30 LeuAlaGlyCysAlaSerIys-----ProThrIYAsnSerThSergIy 44
DB 215187 CTGGCAGGTTGTATCCAGTTCCTCAAACTCCGCGCGGTCACGTCGTGACAGCGTGG 215128
QY 45 SerGlySerHisArgThSergIy-----SergIyGlyLeu 56
DB 215127 TCGTCATCAATATCAATTCGGGATGTGTATCAACGCCGCCAANAATGGCGCGAGC 215068
QY 57 -----AlaIleGlySer 60
DB 215067 CCGCAACAAGCCGCGCAATTCACCCGTCAGCGTCCTGTACGACGACCCATGCAACT 215008
QY 61 GlnValIleThrAspSergIa----- 67
DB 215007 CAACCGGACGAGGAACAACCCGTGCATGGAATAATGGCGAATCGTATACATCGCCAG 214948
QY 68 -----GlyValPro-----AsnArgTYrGlnValIlyGlnIy 78
DB 214947 TATGGGAATATTCGAAAGGTATACGAGGCGGCGGCACTTACACCGTGAATAAGGT 214888
QY 79 AspThrValSerIysIleAlaGlnArgTYrGlyLeuAsnTPARGGluIleGlyHisIle 98
DB 214887 GATACGCTTTTATACATCGCTGATTCACGGAATATTCGCGGATCTGGCCAGAGA 214828
QY 99 AsnAsnLeuAsnSerSeryThrIleTYrThrGlyGlnTPleu----- 113

```

Db 214827 AACGATATTTCTGCCCCGATAGCTGATGAGGCAAACTTTGCAAGTTGTAACGCA 214768
 Qy 114 -----ThleutpSerGlyAspLeuValArgGluArgSerIleSerSerGly 130
 Db 214767 TCCGGTACGGCAATTAATCTGCGGAAATGCGATCAACGAGGAGGATGACAGCAAGAGA 214708
 Qy 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProVal--- 149
 Db 214707 GTTGTCACGAGCTGCAAAATTCACCGTTGCAAGTGGTGTCTCAACCAACAATTACG 214648
 Qy 150 -----GlnGlnHisProAlaValGlnLysProThr-----ProPro 161
 Db 214647 TATTCGAGGCGCTCAGGTGAAAGTGTCAACAAATGTTGCGCAACCAAGCTGTCT 214588
 Qy 162 ValValValValLysLysProThrProThrProProValValGlnGlnProAlaProVal 181
 Db 214587 GGGACGTTGTTCACAGCGCTGTAACGCGACCAACGTTAGCAGCAGCAACCAATGCA 214528
 Qy 182 AlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArg 201
 Db 214527 AGCAGTACGTCAACCAAGCGCGCG-----ATTCCGCAATGCGCG 214489
 Qy 202 TyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGly 221
 Db 214488 TGGCCG-----ACGATGCGCAAGTGTATGCAAACTTTGGC---GCTTCGCAAGCGCGC 214438
 Qy 222 SerThrValThrSerArgLysMetTyrPheSerGlyValArgAspGlyAspLeuIleAsnAla 241
 Db 214437 -----AATTAAGGAGATCGACATTGACAGCGAGTAAAGGACAGCGCTATCGTCGA 214390
 Qy 242 SerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer-----Ile 259
 Db 214389 ACCGCGATGGCGCGCTGTAATGCGCGTAACGCAATGCGTGTACGTAATCTTAAT 214330
 Qy 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnVal 279
 Db 214329 ATCATCAACATCAACATGATTCCTGAGTGCCTACCGCCCATATGATGATGATGATGAT 214270
 Qy 280 LysThrGlyAspThrValArgThrGlyGlnArgIleLysSerMetLysAsnGlnProSer 299
 Db 214269 CCGGAACACAGAGAGTAAAGGCGGAGCAAAATCGCTACTAGTGGTACACCGGAC 214210
 Qy 300 GlyAlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyrValAspProLeuThr 319
 Db 214209 AGCTCAACACGCTGCAATTTGAATTGTTACAGAGGAAATCCGTAACCCGCTCGT 214150
 Qy 320 ValLeu 321
 Db 214149 TATTTA 214144
 RESULT 34
 SDU6131
 LOCUS SDU6131 1290 bp DNA linear BCT 24-FEB-1999
 DEFINITION Salmonella dublin nlpd gene, strain Lane.
 ACCESSION AJ006131
 VERSION AJ006131.1 GI:3150245
 KEYWORDS nlpd gene.
 SOURCE Salmonella enterica subsp. enterica serovar Dublin.
 ORGANISM Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
 REFERENCE 1 (bases 1 to 1290)
 AUTHORS Paesold, G.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAY-1998) Paesold G., Institute of Microbiology,
 Swiss Federal Institute of Technology, Schmelzbergstrasse 7, ZH,
 8092, SWITZERLAND
 2 (bases 1 to 1290)
 AUTHORS Paesold, G. and Krause, M.
 TITLE Analysis of rpos mRNA in Salmonella dublin: identification of
 multiple transcripts with growth-phase-dependent variation in
 transcript stability
 JOURNAL J. Bacteriol. 181 (4), 1264-1268 (1999)

MEDLINE 99138748
 PUBMED 9973354
 FEATURES
 source
 Location/Qualifiers
 1..1290
 /organism="Salmonella enterica subsp. enterica serovar
 Dublin"
 /strain="lane"
 /db_xref="taxon:98360"
 120..1253
 /gene="nlpd"
 120..1253
 /gene="nlpd"
 /codon_start=1
 /protein_id="CA006881.1"
 /db_xref="GI:3150246"
 /translation="MSAGSPKFTVSRIALSLVSLMAGCTSSNPAPVTSVDSGS
 SNTSGMLITTPPKGATTOQPOAPQIOPVQRPVTPQTPVTEOPVOMENGRIV
 YNRQYNIPIKGSYTGSTYTKKGDILFYIATWGDPRDLAQRNLSIAPSLNVCOT
 LQVNASGMPITGNAITQADAAQGVNTRSAQNSIVAVASOPTIYSBSSGSANK
 MLEPNKPAQTVVTAPTVAPTVSTPBNASSTSDAPISARMPRTDQVIENPQASBG
 NKSIDYASKGAIVATADGRVYAGNALRGYNLIILKINDVLSAYAHNDTMLVRE
 QDEVKAGQKIATGSGTGSTRLHFIRKKGKSVNPLRYLPOR"
 BASE COUNT 344 a 326 c 335 g 285 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.000432 Length: 1290
 Score: 209.00 Matches: 84
 Percent Similarity: 38.25% Conservative: 56
 Best Local Similarity: 22.95% Mismatches: 138
 Query Match: 12.51% Indels: 88
 DB: 1 Gaps: 13
 US-10-018-706-2 (1-322) x SDU6131 (1-1290)
 Qy 30 LeuAlaGlyCysAlaSerLys-----ProThrTyrAsnSerThrSerGly 44
 Db 186 CTGGCAGGTTGTACCACTGCTGTAACACCTCCGCGCGCTGACGTTGTGACAGTGGG 245
 Qy 45 SerGlySerHisArgThrSerGly-----SerGlyGlyLeu 56
 Db 246 TCGTCATCAAAATCAACATTCGGGATGTTGATCACACCGCGCAAAATAGGCGGACG 305
 Qy 57 Ala----- 57
 Db 306 ACGCAACAAACCGCGCAACAGCGCGCAAAATCAACCCGTTGACGCTCTGTAACGACG 365
 Qy 58 ---IleGlySerGlnValIleThrAspSerGln----- 67
 Db 366 CCGATGCAAACTAACCGGTCACGAAACAACCCGTCGATGAGAAATGGCGCAATTGTA 425
 Qy 68 -----GlyValPro-----AsnArgTyrGln 74
 Db 426 TACATCGCAGTATGGAAATATTCGAAAGTATGATACGGGTGGCAGTACTTACACC 485
 Qy 75 ValLysGlnGlyAspThrValSerLysIleAlaArgTyrGlyLeuAsnTyrArgGlu 94
 Db 486 GTGAAAAAGGGTATACGCTTTTTCATCGCTGGAATTAACCGGAAACGATTTCCCGGAT 545
 Qy 95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTyrLeu--- 113
 Db 546 CTGGCCCAAGAAACAGTATTTCTGCCCCGATAGCTGATGAGGCAAACTTTGCAA 605
 Qy 114 -----ThleutpSerGlyAspLeuValArgGluArgSer 126
 Db 606 GTTGTAACGATCCGATATGCAATTAATCTGCGGAAATGCGATCAACGAGGAGTGA 665
 Qy 127 IleSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArg 146
 Db 666 GCACAGCAAGAGTGTGACAGGCTGCAAAATTCACCGTTGCAAGTGGTGTCTCA 725
 Qy 147 ProProVal-----GlnGlnHisProAlaValGlnLysProThr 159

Db	1226	CCAAACATTTACGTAATTCGAGGGCTCAGGTGACAAAGTGTCTAAACAAATGTTGCCAAAC	785
Qy	160	-----ProProValValValValValValValValProThrProThrProProValValGlnIln	177
Db	786	AACAAGCTGCTGGGAGCGTTGTACAGCGCTGTAAACGGACCAACGAGTAAGACAGAC	845
Qy	178	ProAlaProValAlaProProValAlaThrGlnAlaProPheAlaThrGlySerSerGlyVal	197
Db	846	GAACCAATAGCAAGCAGTACGTCAACAGCGCGCG-----ATT	884
Qy	198	MetGlnPheAlaGlyProValGlyValAlaThrAsnProValValAlaGlyPheGlyThrAla	217
Db	885	TCCGCATGTCGCTGACCG-----ACGATGGCAAGATGATGAAACTTTGGC---GCT	935
Qy	218	ThrValAlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyValArgAspGlyAsp	237
Db	936	TCCCAAGGGGC-----ATTAAAGGATCGCATTTGCAGCGAGTAAGGAGAG	983
Qy	238	LeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaSerPheAsnMetAspGlyValAla	257
Db	984	GCTATCGTCGCACACCGCATGAGCGCGCTGTATATGCGCGTAACGCACCTCGTGTATAC	1043
Qy	258	Ser-----IleValIleGlnIleThrAsnGlyPheValSerSerTyrlleHisIleIys	275
Db	1044	GGTAATCTTATATCATCAACATCAATCAATGATTTCTCGAGTGCCTTAGCCCATATGAT	1103
Qy	276	AspAlaGlnValIysThrGlyAspPheValArgTrpGlyGlnArgIleAlaSerMetIys	295
Db	1104	ACGATGCTGTCCTCGGAAACAACAGAAATTAAGCGCGGCAAAAATATCGTACTATGGGT	1163
Qy	296	AsnGlnProSerGlyAlaAlaLeuPheGluPheAlaGlyIleSerArgAsnGlyValIleTyra	315
Db	1164	AGCACCGGACACAGCTCTACACGCTTGATTTGAATTTCGTACAAAGGAAATCCGTA	1222
Qy	316	AspProLeuThrValLeu	321
Db	1224	AACCGCTGCGTTATTTA	1241

[illegible][illegible]

RESULT 36					
AE008833/C					
LOCUS		20984 bp	DNA	linear	BCT 31-JUL-2002
DEFINITION	Salmonella typhimurium LT2, section 137 of 220 of the complete genome.				
ACCESSION	AE008833	AE006468			
VERSION	AE008833.1	GI:16421460			
KEYWORDS					
SOURCE					
ORGANISM	Salmonella typhimurium LT2.				
	Salmonella typhimurium LT2				
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.				
REFERENCE	1. (bases 1 to 20984)				
AUTHORS	McClelland, M., Sanderson, K.E., Spiehl, T., Clifton, S.W., Du, F., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, C., Nguyen, C., Scott, K., Holmes, A., Gwena, N., Milvanev, E., Ryan, E., Sun, H., Flores, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.				
TITLE	Complete genome sequence of Salmonella enterica serovar Typhimurium LT2				
JOURNAL	Nature 413 (6858), 852-856 (2001)				
MEDLINE	21534948				
PUBMED	11677609				
REFERENCE	2 (bases 1 to 20984)				

AUTHORS The Salmonella typhimurium Genome Sequencing Project.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER: <http://www.tigr.org/softab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
ReguonDB;
http://kinich.cifn.unam.mx:8850/db/reguondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

FEATURES
source
1..20984
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
71..1003
/gene="STM2912"
71..1003
/note="similar to E. coli putative transcriptional
regulator lysR-type (AAC73313.1); Blastp hit to AAC73313.1
(304 aa), 32% identity in aa 6 - 289"
/codon_start=1
/transl_table=11
/product="putative lysR family transcriptional regulators"
/protein_id="AA121792.1"
/db_xref="GI:16421461"
/translation="MTVKEMLAMLNQKSLFPAVVDGSGPTAAASGQTRAVVFNIR
ROLEKELGVTLLSTRRLTLTDAGVLFYQGVNLLNAKNLQDEVRASHGSGELR
ITTPPEGEVITIPILAQFOSRHPRLIRHSSSHHADLIERPVALRSLADSRV
RALISRRTIIPVAPQVLAAPVUSLESILAOBHTHERLPTPRATVTSNKGCHSR
LEISKAGKISVDSASLSMAFALAGSGVALLPQMLVNTALEBDGTLHVLDPHPRQGI
YAVYPDAHVSTKRAFTDFLRSDQCEHAPSS"
complement(1045..2548)
/gene="STM2913"
complement(1045..2541)
/gene="STM2913"
/note="similar to E. coli putative transport protein
(AAC75782.1); Blastp hit to AAC75782.1 (454 aa), 33%
identity in aa 11 - 446"
/codon_start=1
/transl_table=11
/product="putative permease"
/protein_id="AA121793.1"
/db_xref="GI:16421462"
/translation="MTSLPTPIIGLIVAVFLVWLVLTRVHALIAMLAAACIAGLLG
GMGIDKLTSTGSGTGLGSLGIVIGVAMGRLLEVSAAERLAFIMLGKRRP
EMALITGVIVSIPFVDSAVILVGVKALAKGSKSLTLTGVALAGLAVHTHTYR
PTPGPIAGIIFNDIGAMLLGMAALPCVIGIVPAOMLDRYRDPVPTIADRY
NAALFOYKEKEKESLPSLTSLPVIPIVILFKALCSTLATEGSGATPFRVVO
AINTVGSFVIALAISVLAVYTLVPRMDKHTAELEGLQASGILLVLTAGAGLGA
ILRDSAGQQLAEQVANIIPSPILIPVATLVRLIQSGGTVAMIITASISAPILAQI
PGIMMLAAQATWGSILFPGYFNDLSLFWVVRMKGVSVDKQMVVSVPTTIAAIGG

gene
complement(2538..3495)
/gene="STM2914"
complement(2538..3488)
/gene="STM2914"
/note="similar to E. coli UDP-galactose-4-epimerase
(AAC73846.1); Blastp hit to AAC73846.1 (338 aa), 25%
identity in aa 1 - 162"
/codon_start=1
/transl_table=11
/product="putative nucleoside-diphosphate-sugar epimerase"
/protein_id="AA121794.1"
/db_xref="GI:16421463"
/translation="MQIITGGGFLGQKLASALLNSLARNELLVLDKMPARLSDS
PRACLEADLPQVLESVITRANSTVYTHLAATYSNHEDEPDGKKNLDITROLLE
ACRQPOKTRFVSSSLAVYGTITPECVTDITLPRSYAQAKACLVNDYTRKG
YVDGLARLPETICVRGKPNRPAASPSAIIREPLOGETTCVPSERLWISSATV
IHNLSLATLPAPGEASINLPGISVWGEMLFTLQAGCGQARDRVTHORDEGEKI
VASWPRIDNORALALGFADRFDDIIFERPDMEGRS"
complement(2543..2548)
/gene="STM2914"
complement(2543..2548)
/gene="STM2914"
/note="putative RBS for STM2913; RegulonDB:STMS1H002970"
complement(3490..3495)
/gene="STM2914"
/note="putative RBS for STM2914; RegulonDB:STMS1H002971"
complement(3528..4318)
/gene="ygbM"
/note="STM2915"
complement(3528..4304)
/gene="ygbM"
/note="similar to E. coli orf, hypothetical protein
(AAC75781.1); Blastp hit to AAC75781.1 (258 aa), 79%
identity in aa 1 - 258"
/codon_start=1
/transl_table=11
/product="putative endonuclease"
/protein_id="AA121795.1"
/db_xref="GI:16421464"
/translation="MPRPANLSMFTPEAPFERPAAAEQFQAVELPYPRDPAE
IKQLSHNLTIALFNTSAGDTAGENGKRALPERHDARADITDLAEYALACEVO
HIMGVVDPDGDARVATPFDINLRVADRRPDKILLEALSPPGVKPLFSSQYO
ALGIAEVDPRNVIQDTEFAQVVDNLSLILEYVGRVHVOIASLPDRHREDDGE
INYPMLPRLPDDVDGIRGCEYOPRNTYDQGLQWPMANR"
complement(4309..4362)
/gene="ygbL"
/note="STM2916"
complement(4309..4347)
/gene="ygbL"
/note="similar to E. coli putative epimerase/aldolase
(AAC75780.1); Blastp hit to AAC75780.1 (212 aa), 85%
identity in aa 1 - 211"
/codon_start=1
/transl_table=11
/product="putative fucose phosphate aldolase"
/protein_id="AA121796.1"
/db_xref="GI:16421465"
/translation="MTILAKBHALREBMRIASFPQRYATSGAGNLSLLPDGNT
LATPTGSLGNDLPQRLSKVDPOGEMLSGKPSKVRPHLALYNNPCCCKAVHLST
WSTALSGICLNDPQNVIRPPTPYVVMWGIPVLPYRPGDRRIARLAAALAAHQA
FLAHGVVCGENIQEAAANTELEEFKALIFIGERDIRVLTETBIAQNR"
complement(4313..4318)
/gene="ygbL"
/note="putative RBS for ygbM; RegulonDB:STMS1H002972"
complement(4944..6217)
/gene="ygbK"
/note="STM2917"
complement(4944..6206)
/gene="ygbK"
/note="similar to E. coli orf, hypothetical protein
(AAC75779.1); Blastp hit to AAC75779.1 (388 aa), 79%
identity in aa 1 - 384"
/codon_start=1
/transl_table=11

```

/product="putative RNA synthase"
/protein_id="AAL21797.1"
/db_xref="GI:16421466"
/translation="MLKIGVADDFGTGATDIAFLVENGMPYVOINDPTGTOEGCD
AVVITSLKTRSPGAEAIKQSLALVWLKQCOQVFPYCTPSTAEKNGIPYTDAL
MVALDSTFTVSPALPVNGSTLVYGYLFVNMHLAESGMHRHPINPTDSYLPRLME
OAGRCGVTPACOTIDKGVATRAALSLQSGRYAVLDLNERHLEIOGVLEADPL
VWGGSGCGLMGAROMAHGVSOASAGVPIPSGRAVIVSGSCOMTNOOAFYRQHAP
RDVAVARCSERREAYEAALQWVLSDPSLAIPMISTOLAALQOQYKTEAS
HAYVALPSLAARLARAGITRPFVIAQVCEPSTVOSIGITGFHIGPCISBPVVMNL
HAPVSLAKSNGRDESFTRAGREFV"
/complement(4957..4962)
/gene="ygbL"
/note="putative RBS for ygbL; RegulonDB:STM1H002973"
/complement(6200..7134)
/gene="STM2918"
/note="ygbJ"
/complement(6200..7123)
/gene="ygbJ"
/note="similar to E. coli putative dehydrogenase
(AAC75778.1); Blastp hit to AAC75778.1 (302 aa), 82%
identity in aa 1 - 299"
/codon_start=1
/transl_table=1
/product="3-hydroxyisobutyrate dehydrogenase"
/protein_id="AAL21798.1"
/db_xref="GI:16421467"
/translation="MTGTDFHVGIVGLSGMGMGAARSCLRAGLSTMGADINPOCAN
LLARAGCAASAREFAGVVDVLLVYVNAQVQEVTFGEQVHMLAEKPAVNVST
ISSADQEIATLALTNINMLDAPVSGAVVAAQSEVMAVSGSAAPTRFLKPYLDV
ASNYRISDPTGAGSTYKTIHQLAGVHIAAABEMALAAAGAPLDVWYVYTHAG
NSMPEFMQHVVDYTPRRAVDLFFVMDLGVADTARALKRPLPLASTALMTSAS

```

Alignment Scores:

```

Pred. No.: 0.0038      Length: 20984
Score: 209.00         Matches: 84
Percent Similarity: 38.25%      Conservative: 56
Best Local Similarity: 22.95%    Mismatches: 138
Query Match: 12.51%             Indels: 88
DB: 1                         Gaps: 13

```

US-10-018-706-2 (1-322) x AE008833 (1-20984)

```

QY 30 LeuAlaGlyCyAlaSerIys-----ProThrTyAsnSerThrseryI 44
DB 13112 CTGGCAGGTTGTACCAACCTCTCCGCGCGGTCACGCTGTGACAGTGGT 13053
QY 45 SerGlySerHisArgThrseryI-----SerGlyGlyLeu 56
DB 13052 TCGTCATCAATACCAATTCGGGAATGTGATCACACCGCGCAAAATGGCGGACG 12993
QY 57 Ala----- 57
DB 12992 AGCGCAACAACCGCGGACACCGCGCAATTCACCGCTTCAAGCTCTGTAACGCG 12933
QY 58 ---11eGlySerGlnValIleThrsPserGln----- 67
DB 12932 CCCATGCCAACTCAACCGGTCAGGACCAACCGCTGCAATGGAATAATGGCGAATCGTA 12873
QY 68 -----GlyValPro-----AsnArgTyrGln 74
DB 12872 TACCAATGCCAGTATGGGAATTTCCGAAAGTAGCTATACGGGTGGCAGTACTTAACAC 12813
QY 75 ValIlyGlnGlyAspThrValSerIysIleAlaGlnArgTyrGlyLeuAsnArgIle 94
DB 12812 GTGAAAAGAGGTATACGCTTTTTCATCGCTGATTAACCGGGAACGATTTCCCGCAT 12753
QY 95 11eGlyHisIleAsnAsnLeuAsnSerIysThrIleTyrThrsGlyGlnTrpLeu--- 113
DB 12752 CTGGCCCAAGAAACAGTATTTCTGCCCGGTATAGCTGAATGTAAGGCAAACTTTGCAA 12693
QY 114 -----ThrLeuTrpSerGlyAspLeuIysValArgGluValGlySer 126

```

```

DB 12692 GTTGGTAACCGATCCGTAACGCAATTAATCTGCGGAAATGCGATCACCCAGCGGATGCA 12633
QY 127 11eSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArg 146
DB 12632 GCACAGCAAGAGATTTGGACCAAGTCTGCAAAATTCACCGCTTGCAAGTTGGCTTCAA 12573
QY 147 ProProVal-----GlnGlnHisProAlaValGlnIysProThr 159
DB 12572 CCAACATTCAGTATTCGAGGCTCAGGTGAAACAAAGTCTTAACAAATGTGTCCAAAC 12513
QY 160 -----ProProValValIleValIlyIysProThrProThrProProValValIleGln 177
DB 12512 AACAGCTGCTGGACGTTGTTCACAGCGCTGTAAACGCAACCAACGATACACAGACC 12453
QY 178 ProAlaProValAlaProProValThrsGluAlaProPhaAlaThrIlySerSerIyVal 197
DB 12452 GAACCAATGCACAGTACGTACACAGCGCCCG-----ATT 12414
QY 198 MetGlnPheArgTyrProValGlyAlaThrsAsnProValValArgPheGlyThrAla 217
DB 12413 TCCGATGGCGCTGGCCG-----ACGATGCAAGTGAATCCAAAATTGGC---GCT 12363
QY 218 ThrValAlaGlySerThrValThrsAsnGlyMetTrpPheSerGlyArgAspGlyAsp 237
DB 12362 TCCGACAGGGGCG-----AATAAAGGATTCGACATTGACGCACTAGTAAGGACAG 12315
QY 238 LeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAla 257
DB 12314 GCTATCGTCGACACCGCTGATGGCGCGCTGTATATCCCGGTAAACGACCTGGCTGTAC 12255
QY 258 Ser-----11eValIleGlnHisThrAsnGlyPheValSerIysIleHisIleIys 275
DB 12254 GGTAACTTATTCATCAACCAATACGATGATTAACGTGAGTCTTACCGCCCAATATGAT 12195
QY 276 AspAlaGlnValIlySerThrsGlyAspThrValArgThrGlyGlnArgIleAlaSerMetIys 295
DB 12194 ACGATGCTGCTGCGGGAACAACAGAAATTAAAGCGGGGCAAAAATCGCTATATGGGT 12135
QY 296 AsnGlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyrVal 315
DB 12134 AGCACCGGACACCTTACACGCTTGCAATTTTAATTCGTTACAAAGGGAATCCGTA 12075
QY 316 AspProLeuThrValLeu 321
DB 12074 AACCGGCTGCGTATTTA 12057
RESULT 37
AE004139 10104 bp DNA linear BCT 31-JUL-2000
LOCUS 10104139
DEFINITION Vibrio cholerae chromosome I, section 47 of 251 of the complete
ACCESSION AE004139
VERSION AE004139.1
KEYWORDS GI:9654953
SOURCE
ORGANISM
Vibrio cholerae.
Vibrio cholerae; gamma subdivision; Vibrionaceae; Vibrio.
1 (bases 1 to 10104)
Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L.,
Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Umayam, L.,
Gill, S. R., Nelson, K. E., Read, T. D., Tettelin, H., Richardson, D.,
Ernst, M. A., M. D., Vamathevan, J., Bess, S., Olt, H., Dargatz, I.,
Sellers, P., McDonald, L., Ueberlack, T., Fleischmann, R. D.,
Nierman, W. C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL Nature 406 (6795), 477-483 (2000)
MEDLINE 20406833
PUBMED 10952301
2 (bases 1 to 10104)
Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L.,
Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Umayam, L. A.,
Gill, S. R., Nelson, K. E., Read, T. D., Tettelin, H., Richardson, D.,

```

Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
 Seliars, P., McDonald, L., Utterberg, S.T., Fleischmann, R.D.,
 Niernan, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
 Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
 Direct Submission
 Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 10104
 /organism="Vibrio cholerae"
 /strain="N16961"
 /serotype="O1"
 /db_xref="taxon:666"
 /chromosome="1"
 /note="biotype: El Tor"
 85. .375
 /gene="VC0527"
 85. .375
 /gene="VC0527"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AA093695.1"
 /db_xref="GI:9654954"
 /translation="MWRVFPALTLTLWLLTYLMKNGKVMDFRAVOAIEVQOQV
 NANTLNQMFALIDLRQGLDAIEERANELGMVDFEFYIIIEBSNQ"
 372. 1070
 /gene="VC0528"
 372. 1070
 /gene="VC0528"
 /note="similar to GB:142023 PID:100557 PID:1220756
 PID:1204923 PID:1573673; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="CDP-ribitol pyrophosphorylase-related protein"
 /protein_id="AA093696.1"
 /db_xref="GI:9654955"
 /translation="MNNNTALVPAAGVSGRMQADPRKOYLTLDKTYLERTVEHLEH
 PLIEVVAVASADPYFANLPLAHPRVIRVDGKERADSVLSLEYVCGHRSFWL
 VHDARPCVTHADITLITLTAHPIGAILASPYRDTMKGDHLQOIVHTVQALWHL
 ALTPQMRQASIRERLPAALQOQVITDEASAFEMRGEKPALVAGRADNLIKTPEDL
 ALAEFYLSRNKXKS"
 1067. 1543
 /gene="VC0529"
 1067. 1543
 /gene="VC0529"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AA093697.1"
 /db_xref="GI:9654956"
 /translation="MIRIGGFVDHFRGEGPIIIGVKIPEQGLIAHSDGVATLA
 LSDLIGALIAAGDIGHRPPTDDMKAGDSRELLKDYRRYKAGVYLGNADVTIIO
 APKAPYIOAMCAIADLETDLGNINVKATTEKLGFTGRKEGIAEAVVLRKA"
 1573. 2658
 /gene="VC0530"
 1573. 2658
 /gene="VC0530"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AA093698.1"
 /db_xref="GI:9654957"
 /translation="MTDILSLAYLGGKPTAKAKLALBEHFOVNEVLYGSLGHEH
 LMPVIRKGTGNTSPVANEIAKAGVPSRAVFTGEMLSVHLPNGTDPF
 SAIQAQPSIEHLEVTNRHDKLRGDIQAGNEFWYTLSEVSDVAVALSRLEVAFGLP
 NYFSQRFGRHGNLSERRRGRNVSRNQNOSTLYLSARSVTFNOIVSKRLEQCC
 PARIEDIALAEQOMVDGDLALMDQRLQAGEVALSALAGDNALPTSGOALPLEQ
 AELDAEPDLMALIRGNRMHRDRALAKAQLSWQVEDQITLRFSLDAGSPATSLVR

ELIEIPVNRHYDQGHQESRDESEB"
 2621. 3412
 /gene="VC0531"
 2621. 3412
 /gene="VC0531"
 /note="similar to GB:U00096 PID:1789101 PID:882637;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="survival protein Sure"
 /protein_id="AA093699.1"
 /db_xref="GI:9654958"
 /translation="MNNKRAIIPAKIKMKILISNDGVYAQGIHALADALDAIEVI
 VAPDRNSGASNSLTLEHPLRVQIADENTYSVGTPTDCYHPLNELMKDALPDLVS
 GINHGANGDDVYSGTVAAAMEGHFLGVOSTIPLSAGTTHPASAARFVQIVQHLA
 NPIPTNRLNWNTPDRPLELQIGIVTRLDGRHRSMTIQKOPRGNDIYMLPGKE
 QDAGPTDFPAIERGWSLTPLQVDLTAHSLNSMDHWEKXNG"
 3405. 4031
 /gene="VC0532"
 3405. 4031
 /gene="VC0532"
 /note="similar to GB:M63493 SP:P24206 PID:147115
 PID:882636 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=1
 /product="protein-L-isoaspartate O-methyltransferase"
 /protein_id="AA093700.1"
 /db_xref="GI:9654959"
 /translation="MANPKADRLIQELTEGGITSPOVLAIALPREFFVAPAMQQA
 YDNNAIPIGQGTISQPIYVAKMTIELALPEKTVLEIGSGVQYVLAIVNHT
 VERIKTLQMDAKRKLKOLDYVNSTGHDGDMQMPARGPDALIVTAAAKVQPSLSD
 QLAEGRMVTPVEDDEQYLKTIKRGQGFSEVAVRFPVLAQDLA"
 4031. 4966
 /gene="VC0533"
 4031. 4966
 /gene="VC0533"
 /note="similar to GB:2897901; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="lipoprotein N1p"
 /protein_id="AA093701.1"
 /db_xref="GI:9654960"
 /translation="WTLYSFRGLILFCSILFECTAPTPAPVSGLGKDYKVERGSR
 GSTYEVKGDLYFLIAYLTDKVDNDLISYNDLAPYTHPGQIKMLPYTPPVAGG
 TGAATVVAASSTASVAKAATATVAGTVSKSNKNVONGSTNSQNLTKDKDPVKG
 DQTKKEVVEPVGKQNVNVAKAKPSDEKIAKMPKGRVILKNFSAGDQCKGIDI
 AGORGAVVATADGTIVYSGNALRGYGNLIIIGNHEVLSAVAHNDQLLAKEGQVQA
 GKRIATMGSSTNSVNLHPEIRYQKSVNKRILP"
 5040. 6047
 /gene="VC0534"
 5040. 6047
 /gene="VC0534"
 /note="similar to GB:2897902; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="RNA polymerase sigma-38 factor"
 /protein_id="AA093702.1"
 /db_xref="GI:9654961"
 /translation="MSVSNVTKVEEPPDEALVLETDALTSDEIVAVAGSARL
 VREPPASAKSLDATOMYLSIEIGFSLTRBEVYIARRLRGDEAARKRMISNRL
 VKISRSYSRGLADLDLIEGNLGLRAVEKDPERKGFSTYATWIRQTERALM
 NQRTIRLPLHVKELNIYLTIRARELSQRDHPTEBEALDELDRDVDTKMLRNE
 RISSVDTPIGDGDKALLDLPDSHADPEFSQDDDIRESLNMDELDELPKQEVIA
 RRFGLGYEPSTLEEVGREINILFRERVOIQVEGLRRRLRIIVKQGINMEALPNVEYD
 N"
 complement (6140. 8728)
 /gene="VC0535"
 complement (6140. 8728)
 /gene="VC0535"
 /note="similar to GB:M64730 SP:P23909 PID:146906

the complete sequence.
AF009547 AB008917
AE009547.1 GI:17983036

Brucella melitensis
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Brucellaceae; Brucella.
1 (bases 1 to 12362)
Delvecchio,V.G., Kapataci,V., Redkar,R.J., Patra,G., Mujer,C.,
Los,T., Ivanova,N., Anderson,I., Bhattacharya,A., Lykidis,A.,
Rezink,G., Jablonksi,L., Larsen,N., D'Souza,M., Bernal,A.,
Mazur,M., Goldsman,E., Selkov,E., Elizer,P.H., Hagius,S.,
O'Callaghan,D., Ietesson,J.-J., Haselkorn,R., Kyridides,N. and
Overbeek,R.
The genome sequence of the facultative intracellular pathogen
Brucella melitensis
Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
11756688
2 (bases 1 to 12362)
Delvecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
Direct Submission
Submitted (13-NOV-2001) Institute of Molecular Biology and
Medicine, University of Scranton, Scranton, PA 18510, USA
3 (bases 1 to 12362)
Elizer,P.H. and Hagius,S.
Direct Submission
Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
Center, 111 Dalmatje Building, Baton Rouge, LA 70803, USA
4 (bases 1 to 12362)
Kapataci,V., Los,T., Ivanova,N., Anderson,I., Bhattacharya,A.,
Lykidis,A., Rezink,G., Jablonksi,L., Larsen,N., D'Souza,M.,
Bernal,A., Mazur,M., Goldsman,E., Selkov,E., Haselkorn,R.,
Kyridides,N. and Overbeek,R.
Direct Submission
Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
Park Drive, IL 60612, USA
5 (bases 1 to 12362)
Ietesson,J.-J.
Direct Submission
Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,
Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
61 rue de Bruxelles, Namur 5000, Belgium
6 (bases 1 to 12362)
O'Callaghan,D.
Direct Submission
Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
Kennedy, Nimes 30900, France

1 12362
/organism="Brucella melitensis"
/strain="16M"
/db_xref="taxon:29459"
/chromosome="1"
complement(108..1563)
/gene="BME11069" .1563)
complement(106..1563)
/gene="BME11069"
/EC_number="5.2.1.8"
/codon_start=1
/transl_table=11
/product="TRIGGER FACTOR, PRIASE"
/protein_id="AAU52250.1"
/db_xref="GI:17983037"
/translation="MTRSEGLMNVETETLNEGLKEIKVVPAGDLEAKLERTAR
GARINGEPGVPPTAHLKMYGKSPMAEIVNEIINDSRSILARNESKSAOPVIM
SDEKEAEKVLGKADPVLNVEVLPALEVDPSKIAVTRREVDSIDBEVDQVRI
ASTRTPEFKKGAENEDRVITIDYGLKDGEPFEGADNDQVLVSGQPIFGFEOL
ILKAGDEKIVITVTPPAEYGAHLAKGKATPDIKYEVAKPNELVLDDETKAGIES
LERLRVVEIOESYOYGOITROKVKROIIDLADGDYOFETPOKLVDAEFNIMWQINF
DIQAGRTEDEETTEEARAREYRKLAERVRVLGVLISIKAGVIEVEELQAVY
DQVRVPGQEKIYDPIRTPTPAVNILRIPIREKQVHLLANINVTDKVSKELTA
EBDASAEKPAKKAARAKKAPKKAEGSGEBA"

gene
CDS
complement(1627..1709)
/product="tRNA-Leu"
complement(1943..2512)
/gene="BME11070"
complement(1943..2512)
/gene="BME11070"
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="AAU52251.1"
/db_xref="GI:17983038"
/translation="MTTNKRWLYAGAVLAAALLOTGILVAGIEKRAILRSQGDIVPT
EPVDPDLARGDGVVIVGYEISINARSAGIVPAGSRVYVAVKPAKIGSRPSASF
VPEKDLATGRTGAGYGISDPPANRILSPGIERVYVPEEGHPIDSGQREHDIT
VIVAYDASGPIVKALMDGQPIREGPL"
complement(2509..3513)
/gene="BME11071"
complement(2509..3513)
/gene="BME11071"
/codon_start=1
/transl_table=11
/product="Hypothetical Membrane Spanning Protein"
/protein_id="AAU52252.1"
/db_xref="GI:17983039"
/translation="MATLGGLLGAIVNLVAANQEMRLNRIGVIFVITWASTYCG
AMROAGDKVFPALVYLGASFGAGIALVGMQYHISGDVYSALVYITGLVASFLA
RAQALAFGAGVACFYLTSPADSNLSGADISRYVAGPLLLAGVAAALFTSRRAA
HFLALFSIGCLLVAGQENKTVLLMTIVIGLILADGIRHQLQKTRFPAHPLAY
GLLVLISFALILDVITVGVASGIDRIDILYSMLILALISGAIAICGRDNGILRSI
AYAFSIEVLYLAFTVGTWIGTSGFELTNGILVLLAAFRVRRMESRFRKQGLEAHP
" 3919..4065
/gene="BME11072"
3919..4065
/gene="BME11072"
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="AAU52253.1"
/db_xref="GI:17983040"
/translation="MNLFRSYNWRKRYETVNLNQLSTRELDNGISRADIPVANO
AKAR" 4271..5707
/gene="BME11073"
4271..5707
/gene="BME11073"
/codon_start=1
/transl_table=11
/product="GLUCOSE INHIBITED DIVISION PROTEIN A"
/protein_id="AAU52254.1"
/db_xref="GI:17983041"
/translation="MKALIKDGKIAMSNNTDLSPPVIVIGGLAGSEANQIAGV
VHLHEMPVRGTDAHKTQCLAEVCSNPSFSDAETNAVGLHAKMLAASLIACAD
AHQVPAGALAVDEGFSQAVTARLEAHPITIEREETITIPTEWGTIIITATPLA
PSLAELAAETDADALAFDAIAPIIHFDISINDVCMFQSRVGVGGGKQKINCP
MDKQYAPFAVALIEGDKTDPKEWEGPGLPIVMARERGPETIRHGMKMGILT
NAHNPVYKPAVVOILRODNLAGTLNNVWGFGTKLYSGQGIIFMIGLNAEFAKLG
GLHNTYLSFVLIDNLVRLRSKOTLPAFCQVTCSEGVYSSAIGLAGRTAEKLS
QAAYPPPTTAFGLIGHTIGGHI VTDDDEGKGSFGPMNVNFGLPFPVDVYKPKGKEL
RGEKTIKKRAKLSARALADCRNMLSLY"
complement(15759..6592)
/gene="BME11074"
complement(15759..6592)
/gene="BME11074"
/EC_number="2.5.1.-"
/codon_start=1
/transl_table=11
/product="PHYSIOME SYNTHASE"
/protein_id="AAU52255.1"
/db_xref="GI:17983042"
/translation="MNQETHCLALLREADRRLYSVYAPQERKALAAIYARNAEI
ARIDLVIEPPLGQVRLOMWRDLINGEARGSAEAPAAALIEITIEKLEPIAPDN
CEARVFDLYDDPMPSRNDLEGYCGETASALIQLAGFIIDRDAAQAAHETAGHAGVAA

VAGLLRLPLHRRRGQVVPADMLKAIGVTRDMLSSSEKATERAVTIMALAREHL
 AIFPKAKAGLPOTLHPAFLPMALVPALNAVERIGAQAQVADISAIRKQMLFRAN
 F"
 complement (6598) .6987)
 /gene="BME11075"
 /complement (6598) .6987)
 /gene="BME11075"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical Cytosolic Protein"
 /protein_id="AAL52255.1"
 /db_xref="GI:17983045"
 /translation="WAKGIEIRAEHPGRADIDAYNGGFRFAEMSHRSILCLPSGI
 HGMKPNPPLSRADIGALEEASDIEILVGTGMDLRRIPEDEVALLRHHISSDPM
 STGAAYRYVNLAEADRAVAALAVD"
 complement (6990) .9317)
 /gene="BME11076"
 /complement (6990) .9317)
 /gene="BME11076"
 /codon_start=1
 /transl_table=11
 /product="PROTEIN TRANSLOCASE SUBUNIT SECD / PROTEIN
 TRANSLOCASE SUBUNIT SECF"
 /protein_id="AAL52257.1"
 /db_xref="GI:17983044"
 /translation="WGLRQCFIOAETDLHSMLEYFRKWSALIMLAVLSLIPSPN
 LSRETLNLPDLPRKQVSLGLSGSRLLIYONMGKTDLETTANIMRRLBELY
 GNIVYBERGNOQIRVEVDGLYDAQLMDILIRNLSGRANDMSDDARQGPLAD
 SEIYSGDDPVPGLTKETPLTGHADITDAKSTISADGGQVITLTLDDNRRLADI
 TAQGENSFALVNDQVASFVSGPLDSELIQEGADQVANNMVLKSLPDA
 VTYLERTIASALGEDYASAVLALALAVGLPMVLSYGLIIVLAVNVIITL
 TAVSLIGASISLAIAGLVLTIGLADAHILYERVEDRKGYSVQAMESEFYA
 LSTIVDNLTLTIALVALFLISGTVGPAFLVAGITLFTTLTFTLLIAGMVRT
 AKPEVEKRLKIVPTVTHIDPMRLQVPTTIGISLACIYVALFVNIGYNDIRGS
 SMVELQKNGDANLEEDINERLAEINDSARVLPKSPRSALVITGSGVEDDAQITVA
 VKLSTHEDVLSGMEIVFRMFNLSVAALITIGISLNDVIVYDRENLREYKKA
 PLPAIDIASINQTSRLTSTFTFLAHPVLAQVLAAGSGSEIRMFALASGITIVASYSI
 FLAFLVQFLKRETDAGDAVDALAQSLNLS"
 complement (9437) .9799)
 /gene="BME11077"
 /complement (9437) .9799)
 /gene="BME11077"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical Exported Protein"
 /protein_id="AAL52258.1"
 /db_xref="GI:17983045"
 /translation="MTSEEFPLMFTPAFAQASGVGPDMLMSILPFLIFVNYFLI
 IRPORTOKKROEMLSYVRGDIYVTGGGIVGKVLKQVDDNELELELADGVRIHVRA
 TLMDVRVKGEPVANRKNK"
 9902 .10804

gene

Alignment Scores:

Pred. No.: 0.00902 Length: 12362
 Score: 205.00 Matches: 79
 Percent Similarity: 41.06% Conservative: 45
 Best Local Similarity: 26.16% Mismatches: 132
 Query Match: 12.28% Indels: 46
 DB: 1 Gaps: 12

US-10-018-706-2 (1-322) x AB009547 (1-12362)

Qy 30 LeuAlaGlyCysAlaSerLysProThrTyraSerThrSerGlySerGlySerHisArg 49
 Db 11653 ATCGCTGGCTTCGCGACACGCTGCTGTCACACGCAAGAAATCCTCCGACCAATATG 11594
 Qy 50 ThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyVal 69
 Db 11593 GCATCGCTGGCGGCGCTATGCA-----ACCCGCGCATCATG 11555
 Qy 70 ProAsnArgTyrGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTyrGly 89

Db 11554 GCGGTCCTTAAGCTGTGAAGAGCGGGGATTGCTGTTCTCCATCGCAAGAACACAAAT 11495
 Qy 90 LeuAsnTrpArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyThrIleTyThr 109
 Db 11494 GTCCCTGTCAACGATTGAAGCGCGCAACGCGCTTACCACTGCG---GCCATTGCGGTC 11438
 Qy 110 GlyGlnTrpLeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSerSer 129
 Db 11437 GGGCAGTCGCTCGTATT-----CCGTCGTGCGCGCA 11405
 Qy 130 GlyValAsnThr-----AlaHisThrProSerProVal---AlaValGlnSerSer 145
 Db 11404 GGTGCGCGCACACAGTGGCCCGCATACCGCGAGCGCGMAACCGCGCAAGCAAGCC 11345
 Qy 146 ArgProProValGlnGlnIleProAlaValGlnLysProThrProProValValVal 165
 Db 11344 AGTCCGCGCGCAATGAGAGGTGCGAGAGTAAAGCTTATACCGCGCGCGAGCCAGCAAC 11285
 Qy 166 LysLysProThrProThrProProValValGlnGlnProAlaProValAlaProProVal 185
 Db 11284 AAG-----GTATCGAGATGCGGAA----- 11264
 Qy 186 ThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGly 205
 Db 11263 AAGATCAGGCGCGAGCGCCATCCTCAACCGGCATCTCGCAGATGCGTGGCCGTT--- 11207
 Qy 206 AlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThr 225
 Db 11206 ---CGTGGCGCGCATTCGCGCAAGCTTGGCCAGCGTGA---GGCAGCTCGGTC 11159
 Qy 226 SerAsnLysMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGly 245
 Db 11158 AGCAGCGATATCATATCATATCATATCGTCCGGAAGCAGCGCGCAAGCGCGAAGCGGT 11099
 Qy 246 ThrValIleGlnAlaAspHisAsnMetAsp-----GlyAlaSerIleValIleGlnHis 263
 Db 11098 GTTGATGATCATAGCGGCGCATGTTGATGAAGAAATTTGCGCAGCGCTTGATCCGCGAC 11039
 Qy 264 ThrAsnGlyPheValSerSerTyThrIleHisIleLysAspAlaGlnValLysThrGlyAsp 283
 Db 11038 GACATGCGCTTGACGCTGTATGCGCATTAACAGCCAGATATGTGTGACGCGGCGCAG 10979
 Qy 284 ThrValArgThrArgGlnArgGlyLeuAlaSerMetLysAsnGlnProSerGlyAlaAla--- 302
 Db 10978 AAGTTGCGCGTGGCGAGCAATGCG-----AAGTCGCGCATGACCGCGCAAGCCCAAG 10925
 Qy 303 -----LeuPheGluPheArgGlyLeuSerArgAsnGlyValTyraLysProLeuThrVal 320
 Db 10924 TCGCGAAGCTGCACTTCGAGTGCAGAGAACTCGCGCGCGTCATTCGACCAATAT 10865
 Qy 321 LeuLys 322
 Db 10864 CTGGAA 10859

RESULT 40
 NMA522491/c 329861 bp DNA linear BCT 04-DEC-2000
 LOCUS
 DEFINITION
 Neisseria meningitidis serogroup A strain Z2491 complete genome;
 segment 5/7.
 VERSION
 AL162756 AL157959
 AL162756.2 GI:7380091
 KEYWORDS
 Neisseria meningitidis Z2491.
 Neisseria meningitidis Z2491.
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.
 REFERENCE
 1 (bases 1 to 329861)
 Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
 Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
 Davies,R.M., Davis,P., Devlin,K., Fellwell,T., Hamlin,N.,
 Holtroyd,S., Jagsels,K., Leather,S., Moutie,S., Mungall,K.,
 O'Neill,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
 Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.

TITLE Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491
JOURNAL Nature 404 (6777), 502-506 (2000)
MEDLINE 20222556
PUBMED 10761919
REFERENCE 2 (bases 1 to 329861)
AUTHORS Partholl, J.
JOURNAL Direct Submission
Submitted (30-MAR-2000) Submitted on behalf of the *Neisseria* sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT Notes: Details of *N. meningitidis* sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
Location/Qualifiers
1. 329861
/organism="Neisseria meningitidis Z2491"
/strain="Z2491"
/db_xref="taxon:122587"
/note="serogroup: A"
194. 565
/gene="NMA1439"
194. 565
/gene="NMA1439"
/note="NMA1439, possible lipoprotein, len: 123 aa; unknown, contains a probable N-terminal signal sequence and an appropriately positioned PS00013 prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=1
/product="putative lipoprotein"
/protein_id="CAB84675.1"
/db_xref="GI:7380092"
/db_xref="SPTREMBL:O9JU92"
/translation="MAKTLISIPVALILGGCAAGGNPFGSLDGTGKGSIVKAYE
SCRRBLNRSBMRITALMSAKQDMENKICACVADAPQLTGNVDMQMLDSTR
NOALALITAKTYSACFKLYR"
212. 244
/gene="NMA1439"
/note="PS00013 prokaryotic membrane lipoprotein lipid attachment site"
complement(571. 580)
/note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
592. 601
/note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
638. 2047
/gene="chrc"
638. 2047
/gene="chrc"
/EC_number="4.2.99.2"
/note="NMA1440, thrC, probable threonine synthase, len: 485 aa; similar to many e.g. SW:THRC_METG (EMBL:DJ4071), thrC, *Methylobacillus glycoense* threonine synthase (EC 4.2.99.2) (475 aa), fasta scores: E(): 0. 61.9% identity in 475 aa overlap. Contains Pfam match to entry PF00291 S T dehydratase, Pyridoxal-phosphate dependent enzymes"
/codon_start=1
/transl_table=1
/product="putative threonine synthase"
/protein_id="CAB84676.1"
/db_xref="GI:7380093"
/db_xref="SPTREMBL:O9JU91"
/translation="MKYISTRGETAHAKPSSEVLMLGAPDGLMPEHYPOIGREALD
KWRGLAYPELAEIMRLPVTDIPEDLDLIRKYTEAFGKEITPVKTLSDGIKIO
ALNSGPTLAKOMAFLOFAGAEVYLNKGLNITLIGASGDTGSAAEVALKGGKGVN
VMSLPDGKMSAFORAMYSILOGDNININAVKGMFDDCODIVKAVQNDAAFEKXGIG
TVNSINWGRIVAVQVYVYFAGYFKATQSDNDEQVSCVPSNFENVCAGHIAKMGGLPVR
RLIVATNENDIVLDEFFKTAGYRPNASAHYVSSPMISKSNFERFVFADIMRDPQ
EINTLMAEVAAGKGNRLFALDKVGKGFSTGKSTHADRLLTIROYVEODELIDPH
TADGVKVAEREVERETVCTETALAAKADATIRAVGVDAIPRALGLENLIPOR

misc_feature 875. 1780
/gene="chrc"
/note="Pfam match to entry PF00291 S T dehydratase, Pyridoxal-phosphate dependent enzymes, score 231.80, E-value 1e-65"
complement(922. 931)
/note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
2094. 2888
/gene="NMA1441"
2094. 2888
/gene="NMA1441"
/note="NMA1441, len: 264 aa; unknown"
/codon_start=1
/transl_table=1
/product="hypothetical protein NMA1441"
/protein_id="CAB84677.1"
/db_xref="GI:7380094"
/db_xref="SPTREMBL:O9JU90"
/translation="MTVMQETHSILDELAAKQEAALAPHLADGTXYWRKAGEH
NARWRVALGVAVYKLVKLVKVPVSGEPALATESKRLYEALAGIAVPELAVRK
NALMFGNLEGIPLDTQIROBAEAGKADAMVAGLAEIARVHKQOPLSOAFARMMSDG
KNISFLDFEDDPSSEVLTIACQADMDICLHSTALILKNGSLLEAAAEKKGVLSDP
AEIOKLIATVTKPILPIRRLHPRWGDALRLAASISLISLADMP"
2904. 2937
/note="Stem loop containing DNA uptake sequences: acgat gccgctcgaa gccgctcgaa atata"
2909. 2918
/note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
complement(2923. 2932)
/note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
3065. 3068
/gene="fpr"
3076. 3852
/gene="fpr"
3076. 3852
/gene="fpr"
/EC_number="1.18.1.2"
/note="NMA1442, fpr, probable ferredoxin-NADP reductase, len: 258 aa; similar to many e.g. SW:FENR_AZOVI (EMBL:U36319), fpr, *Azotobacter vinelandii* ferredoxin-NADP reductase (EC 1.18.1.2) (257 aa), fasta scores: E(): 0. 69.8% identity in 255 aa overlap. Similar to NMA1664, fasta scores: E(): 1.3e-30, 34.6% identity in 257 aa overlap. Contains Pfam match to entry PF00175 oxidored fad, Oxidoreductase FAD/NAD-binding domain"
/codon_start=1
/transl_table=1
/product="putative ferredoxin-NADP reductase"
/protein_id="CAB84678.1"
/db_xref="GI:7380095"
/db_xref="SPTREMBL:O9JRE3"
/translation="MAAFNOKVLSVHMWDAYPTPTCTRDESLRFENGGFVWGLMV
DGRMLAYASVANSWMEHLEFISIKVODGPTSLQLHLKAGDVLISKEPTGLVAG
DINNGKILYLLSTGTGAPLSITKDEIYBOPEKILIVGVKKOLAYDRTKEL
BEHYLDIVKEKILYIPVRSREHREHGLTDLMSVSKFEDIGLPIKNPQDRAWL
CGSARLMDCKVLDKGLVSPRTGVRSYLIERAVVD"
3394. 3766
/gene="fpr"
/note="Pfam match to entry PF00175 oxidored fad, Oxidoreductase FAD/NAD-binding domain, score 20.00, E-value 8.4e-05"
3866. 3875
/note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
3964. 3983
/note=">= 90% match to ATCCNNNNNNNGGAAT"
/label=drk3
complement(3984. 4038)
/note="RS103"
/label=RS103

```

repeat_unit 4039..4058
              /note="> 90% match to ATTCNNNNNNNNGGGAAT"
              /label=DRS3
gene         4159..4320
              /gene="NMA1443"
CDS         4159..4320
              /gene="NMA1443"
              /note="NMA1443, len: 53 aa; unknown, lies within a region
              of unusually low GC content"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein NMA1443"
              /protein_id="CAB84679.1"
              /db_xref="GI:7380096"
              /db_xref="SPTREMBL:Q9JUB9"
              /translation="MIYKNCNPKCKAENCKMKMVFLENGSATVHRCIWKIIT
              IDIITPRA"
              /complement(4509..6980)
              /gene="NMA1444"
              /complement(4509..6980)
              /note="NMA1444"
              /note="NMA1444, probable P-type cation-transporting
              ATPase, len: 823 aa; similar to many e.g. SW: COPA_ENTHR
              (EMBL:U13292), COPA, Enterococcus hirae
              copper/potassium-transporting ATPase A (EC 3.6.1.36) (727
              aa), fasta scores; E(): 0, 34.0% identity in 744 aa
              overlap. Similar to NMA1539, fasta scores; E(): 0, 34.4%
              identity in 735 aa overlap. Contains hydrophobic, probable
              membrane-spanning regions. Contains two Pfam matches to
              entry PF00122 E1-E2_ATPase, E1-E2_ATPases and P500154
              E1-E2_ATPases phosphorylation site"
              /codon_start=1
              /transl_table=11
              /product="putative P-type cation-transporting ATPase"
              /protein_id="CAB84680.1"
              /db_xref="GI:7380097"
              /db_xref="SPTREMBL:Q9JUB8"

```

Alignment Scores:

Pred. No.:	0.444	Length:	329861
Score:	203.00	Matches:	71
Percent Similarity:	43.46%	Conservative:	42
Best Local Similarity:	27.31%	Mismatches:	97
Query Match:	12.16%	Indels:	50
DB:	1	Gaps:	11

US-10-018-706-2 (1-322) x NMA522491 (1-329861)

```

Qy 73 TyrGlnVallybGnglyAspThrValserlysllealaglnArglyLeuAsnTrp 92
Db 283427 CATACTATGTGCGCGCGGACCGGTGACCAACATTTCGAACGTCATATCTCTCAA 283368
Qy 93 ArgGluilegIYHsileAsnAsnleuAsnserSerlyThrileYThrGlyGlnTrp 112
Db 283367 GACGATTTCCGTGCGTGAACGCGATGACCGACAAT--ACGTGAGCATCGTCAAG--- 283314
Qy 113 LeuThrleuTrpSerGlyAspLeuIyValArgGluArgserIleserSerGlyValAsn 132
Db 283313 -----ATTGTTAAAGTCAACCGGACGAGATATGCGCA----- 283281
Qy 133 ThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152
Db 283280 -----CCGAAGCCGCGACCGCTTAAA--ACGAGCGCCGCGGTACCG----- 283242
Qy 153 ProAlaValGlnIyLysProThrProProValValValValValValValValValVal 172
Db 283241 GCTGCGGTGCAACCCCTGTG-----AAACCGCGCGCGCAACCG 283203
Qy 173 ProValValGlnGlnProAlaProValAlaPro----- 183
Db 283202 CCTGTGCGAGTCCGCGCGCAACCTGCGCGCCGCTGCGGAAATAAAGCGGTCCGCG 283143
Qy 184 ProValThrGluAlaProPheAla-----ThrGlySerSerGlyValMetGln 199

```

```

Db 283142 CCCCCCGCAATCTCTGCGCTTCCCTCCGCGACGGGTTCGTGCGCGCATTTGT 283083
Qy 200 PheArgThrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrVal 219
Db 283082 TGGCAGCGTCCG-----ACGCAAGTAAAGTGTTCCTCCATTTCCGCGCGC----- 283038
Qy 220 AlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIle 239
Db 283037 -----AACACAAAGCGTGTGATTTTCAGGAATAAGGGGACAGCCGTT 282993
Qy 240 AsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer-- 258
Db 282992 TTGCGCGCGCTGACGCGCAAGTGCTTTATGACAGTTCCGGTTGAGGGGATACGCCAAT 282933
Qy 259 ---IleValIleGlnHisThrAsnGlyPheValSerSerlyrIleHisIleIyAspAla 277
Db 282932 TTGGTCATCATCCGCAATTAATTCCTTCCTGACCGCATACGGCAACCAAAATTTG 282873
Qy 278 GlnValIySerThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetIyAsnGln 297
Db 282872 CTGCTCCGCGCAAGCCACAGGTCAACCGCGGACGACAGTCTTGATGCGCAATACC 282813
Qy 298 ProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAsnGlyValIyValAspPro 317
Db 282812 GAGGCTTCTAGAACGACGATTCATTTCAGAGTGGCGCAAAACGCAACCGTTAATCCG 282753

```

Search completed: July 3, 2003, 21:18:47
Job time : 2677 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 3, 2003, 13:40:17 ; Search time 198 Seconds

(without alignments)
3662.341 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670

Sequence: 1 MTWTAINSONQKPIKRLGL.....LFEFRISRNGVYDPLTVLK 322

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Dgapop 6.0 , Dgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Command line parameters:

-MODE=frame.p2n.model -DEV=xlh
-Q=/cgru2_1/USPTO.spool/US10018706/runat.30062003.091104.23784/app_query.fasta_1.519
-DB=N.Geneseq.101002 -OPMT=faetap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -NORM=ext -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MTLEN=0 -MAXLEN=200000000
-USER=US10018706@cgru2_1.1.200@runat.30062003.091104.23784 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.101002.*

1: /SIDS2/gcgdata/geneq/geneq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneq/geneq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneq/geneq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneq/geneq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneq/geneq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneq/geneq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneq/geneq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneq/geneq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneq/geneq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneq/geneq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneq/geneq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneq/geneq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneq/geneq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneq/geneq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneq/geneq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneq/geneq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneq/geneq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneq/geneq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneq/geneq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneq/geneq-emb1/NA2000.DAT.*
21: /SIDS2/gcgdata/geneq/geneq-emb1/NA2001.DAT.*
22: /SIDS2/gcgdata/geneq/geneq-emb1/NA2001B.DAT.*
23: /SIDS2/gcgdata/geneq/geneq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1670	100.0	969	22 AAF30046	Moraxella catarrhalis
2	1670	100.0	48328	22 AAF28540	Genomic fragment #
3	1666	99.8	966	22 AAF30047	Moraxella catarrhalis
4	219.5	13.1	1830121	17 AAF742063	Haemophilus influenzae
5	209	12.5	2179	21 AAF93349	Haemophilus somnus
6	200	12.2	1248	21 AAF53075	Neisseria meningitidis
7	200	12.0	1014	21 AAF53073	Neisseria meningitidis
8	198	11.9	1056	21 AAF53074	Neisseria meningitidis
9	198	11.9	6100	21 AAF81766	Neisseria meningitidis
10	198	11.9	349980	21 AAF21610	Neisseria meningitidis
11	198	11.9	349980	21 AAF21611	Neisseria meningitidis
12	198	11.9	837096	21 AAF81489	N. meningitidis pa
13	165	9.9	1885	14 AAF51084	H. somnus lppb gen
14	158.5	9.5	684707	24 ABO67196	Listeria innocua c
15	158.5	9.5	2944528	24 ABA03041	Listeria monocytogenes
16	158.5	9.5	3011208	24 ABO68245	Listeria innocua d
17	147	8.8	2365589	24 ABA90521	Genomic sequence o
18	146.5	8.8	615	22 AAF67706	C glutamicum codin
19	146.5	8.8	349980	22 AAF68532	Actinobacillus act
20	145	8.7	949	22 AAF59421	Polynucleotide seq
21	140	8.4	14063	20 AAX20500	Listeria innocua c
22	139.5	8.4	1163020	24 ABO67197	Listeria innocua c
23	135	8.1	1830121	17 AAT42063	Haemophilus influe
24	134	8.0	1233	24 AAL46624	H influenzae BASB2
25	131	7.8	1233	24 AAL46624	H influenzae BASB2
26	131	7.8	3646	14 AAO51086	Plasmid PCR82 whi
27	130.5	7.8	1929	17 AAT06135	lysin coding seque
28	130.5	7.8	4093	21 AAF23454	CDNA encoding huma
29	130.5	7.8	11115	23 ABL50562	Micromonospora car
30	130.5	7.8	15872	18 AAT68715	Streptomyces venez
31	130.5	7.8	15872	21 AAZ87283	S. venezuelae vep
32	130.5	7.8	2365589	24 ABA90521	Genomic sequence o
33	130.5	7.8	4403765	22 AAF96882	Mycobacterium tube
34	130.5	7.8	4411529	22 AAF96882	Mycobacterium tube
35	130	7.7	1233	24 AAL46625	H influenzae BASB2
36	128.5	7.7	1520	21 AAF49721	Achromobacter lyti
37	128.5	7.7	111309	20 AAX20250	Borrelia burgdorfe
38	128.5	7.7	910715	20 AAX20248	Borrelia burgdorfe
39	127.5	7.6	1520	13 AAO25083	Beta-lytic protease
40	127	7.6	7277	20 AAF20026	HEV-US2 full lengt
41	127	7.6	7277	20 AAF20026	Arabidopsis thalia
42	125.5	7.5	942	24 ABR98586	Human cervical can
43	125.5	7.5	1536	22 AAF73004	Nucleic acid seque
44	125	7.5	903	20 AAO00243	Human CDNA seque
45	124.5	7.5	3004	22 AAF15779	

ALIGNMENTS

RESULT 1
ID AAF30046 standard; DNA; 969 BP.

AC AAF30046;

DT 23-APR-2001 (first entry)

XX Moraxella catarrhalis DNA encoding BASB10 protein.

XX BASB10; infection; otitis media; pneumonia; gene therapy;
XX diagnosis; antibacterial; antimicrobial; genetic immunisation;
XX vaccine; ss.

OS Moraxella catarrhalis.

XX Key Location/Qualifiers
FH conflict replace(310,G)

/*tag= a
/note= "gene sequence differs from PCR product
of AAF30047 at this base"

WO200100838-A1.

04-JAN-2001.

23-JUN-2000; 2000WO-EP05854.

25-JUN-1999; 99GB-0015031.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Thomard J;

WPI: 2001-112459/12.

P-PDB; AAB20106.

Novel BASB110 polypeptides of Moraxella catarrhalis, useful as a vaccine for treating Moraxella catarrhalis infections -

Claim 10; Page 81; 88pp; English.

The present sequence is that of the coding region of the BASB110 gene of Moraxella catarrhalis Mc2931 (ATCC 43617), a causative agent of otitis media in children and pneumonia in adults. The gene encodes the 322-amino acid BASB10 protein (see AAB20106). It differs from a BASB110 DNA PCR amplification product (see AAF30047) by having A at position 310 rather than G. This would cause an amino acid change from Ser-104 to Gly in the translated sequence. The invention provides BASB110 polypeptides, and polynucleotides encoding them, as well as expression vectors, host cells and methods for producing BASB110 polypeptides using recombinant methods. Also claimed is a vaccine composition comprising a BASB110 polypeptide, an immunogenic fragment of a BASB110 polypeptide, or a polypeptide having at least 85% amino acid sequence identity to BASB110, or comprising a polynucleotide encoding such a polypeptide. A claimed method of diagnosing a Moraxella infection involves identifying a BASB110 polypeptide or antibody. A claimed therapeutic composition useful in treating humans with M. catarrhalis infection comprises at least 1 antibody directed against a BASB110 polypeptide. CC polynucleotides also have utility in diagnosis of the strage and CC type of infection, and also for therapeutic or prophylactic purposes, in particular genetic immunisation.

Sequence 969 BP; 265 A; 214 C; 242 G; 248 T; 0 other;

Alignment Scores:

Pred. No.:	1-876-116	Length:	969
Score:	1670.00	Matches:	322
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
	22	Gaps:	0

US-10-018-706-2 (1-322) x AAF30046 (1-969)

QY 1 MeThrValThrIleAlaIleAsnSerGlnAsnGlnuysProIleuysArgLeuGlyLeu 20
Db 1 ATGACTGTGAGATGACATCAATTCACAAATCAAAACCATCAAGCATGGCTTG 60
QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40
Db 61 ATTTTGTGTGATGACACCATCTTGTGGCAGATGTGCGATAGCCCACTATAT 120
QY 41 SerThrSerGlySerGlySerHleArgThrSerGlySerGlyGlyLeuAlaIleGlySer 60
Db 121 AGTACCTCAGGTTGGGCGATGATCGTACTTCAGGTTCAGTGGTTTGGCAATAGTTCA 180
QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValGlnGlnGlyAspThr 80
Db 181 CAGGTATCAAGGACATGACAGGCGGTACAAATGCTATACAGTGAAGCAGGGGACT 240

QY 81 ValSerIleAlaGlnIleArgTyrGlyLeuAsnThrArgGluIleGlyHleAsnAsn 100
Db 241 GTCAATGATTTGCTCAGCGTTATGATGATTAATTTGGCTGATGACACATTAAATAT 300
QY 101 LeuAsnSerSerTyrThrIleTyrThrGlyGlnThrLeuThrLeuThrSerGlyAspLeu 120
Db 301 CTAAATGACGATTATACGATTTATACAGGTCATGCGTCACTTATAGTCAAGGATCTC 360
QY 121 LysValArgGluArgSerIleSerSerGlyValAsnThrAlaHleSThrProSerProVal 140
Db 361 AAGGTGGTATGACCTATATACGCTGCTGCTGTAATACGCTCAACACCTTGCGCTGTG 420
QY 141 AlaValGlnSerSerArgProProValGlnGlnHleProAlaValGlnuysProThrPro 160
Db 421 GCGGTTCACTCAGCAGACACACACAGTACAGACATCCTCGCTACAAACCCACGCA 480
QY 161 ProValValValIleValIleValProThrProThrProProValValGlnGlnProAlaPro 180
Db 481 CTTGTTGTGTGTAATAAAACCCACACGACTCCGCTGTGTTGACGACGACGACCA 540
QY 181 ValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
Db 541 GTTGCCCAACAGTACAGACAGACATTTGCCACGGTACCTCAGGGGTGATGCAATTT 600
QY 201 ArgTyrProValGlyValThrAsnProValValArgArgPheGlyThrAlaThrValAla 220
Db 601 CGGTATCTGTGTGTGACCAATCCAGTGTTCGACCTTTGTGACGCGACAGTGGCC 660
QY 221 GlySerThrValThrSerAsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsn 240
Db 661 GCGTCACTTTCCAGTATGACATGCTGTTTTCGACGGATGAGATGATTAATTAAC 720
QY 241 AlSerAsnAlaAlaThrValIleGlnAlaAspHleAsnMetCaspGlyAlaSerIleVal 260
Db 721 GCGGATATGACAGCAGACATTCATCAAGCTGATCAAAATATGACGGGCGAGTATTTG 780
QY 261 IleGlnHleThrAsnGlyPheValSerSerTyrIleHleIleLysAspAlaGlnValIys 280
Db 781 ATTCAGATACCATGATGATTTGTTTCACCTATATTCATATTAAGACCTCAAGTTAAA 840
QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGly 300
Db 841 ACAGGGGATACCGGTGTCACGCTCAGCCTATTGCAAGCATGAAAATCAGCAAGCGGT 900
QY 301 AlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeuThrVal 320
Db 901 GCGGCACTATTGAAATTTGAAATTTCTAGAAATGCGGTATGATGATTCATTGACAGTA 960
QY 321 LeuLys 322
Db 961 CTTAAA 966
RESULT 2
AAF28540
ID AAF28540 standard; DNA; 48328 BP.
XX
AC AAF28540;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #27.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
XX bronchopulmonary; endocarditis; meningitis; ss.
XX Moraxella catarrhalis.
XX
XX WO200078968-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.
 PR (INCY-) INCYTE GENOMICS INC.
 PA Lagace RE, Patterson C, Berg KL;
 PI WPI; 2001-041427/05.
 DR
 XX Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -
 XX
 XX Claim 1; Page 223-235; 545pp; English.
 XX
 XX The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.
 CC
 XX
 SO Sequence 48328 BP; 14211 A; 9322 C; 10802 G; 13992 T; 1 other;
 Alignment Scores:
 Pred. No.: 2.09e-114 Length: 48328
 Score: 1670.00 Matches: 322
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0
 US-10-018-706-2 (1-322) x AAF28540 (1-48328)

DB 41330 CCTGTTGTTGTAATAAACCACACCGACTCCGCTGTTGTTACAGCCAGCACCA 41389
 QY ValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
 DB 41390 GTTGCCCAACAGTGAACAAACACACATTTCACAGGAGTACCTCAGGGGTGATGCAATTT 41449
 QY 201 ArgTyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAla 220
 DB 41450 CGTATCTCTGTTGGGACCAATCCAGTGGTGGAGCTTTGGTATACGGCCAGCACTGGCC 41509
 QY 221 GlySerThrValThrSerSerGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsn 240
 DB 41510 GGCTCACTGTTACAGTAATGGAATGGCATGTGTTCTGTGACAGATGGCGATTTAATTAC 41569
 QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleVal 260
 DB 41570 GCCAGTAATGCAAGGACAGTATTCATCACTATCAATATAGACGGGGCCAGATTTG 41629
 QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleGlyAspAlaGlnValIys 280
 DB 41630 ATTGAGCATACCAATGGATTGTTTCAAGCTATATCCATATTAAGGACGCTCAAGTTAA 41689
 QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetGlyAspGlnProSerGly 300
 DB 41690 ACAGCGCATACGCGTCCGTACCGGTACGGTATTTGCACAGCAAAATCAGCCAAACCGGT 41749
 QY 301 AlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyrValAspProLeuThrVal 320
 DB 41750 GCGGCACTATTGAATTTAGAAATTTCTAGAAATGGCGTGAATGTGATTCATGACAGTA 41809
 QY 321 LeuIys 322
 DB 41810 CTTAAA 41815
 RESULT 3
 ID AAF30047 standard; DNA; 966 BP.
 AAF30047
 AC AAF30047;
 XX 23-APR-2001 (first entry)
 DT
 XX Moraxella catarrhalis DNA encoding BASB110 protein.
 DE
 XX BASB110; infection; otitis media; pneumonia; gene therapy;
 KW diagnosis; antibacterial; antimicrobial; genetic immunisation;
 KW vaccine; ss.
 OS
 XX Moraxella catarrhalis.
 OS
 XX
 FH Key location/Qualifiers
 FT conflict /*tag= a
 FT /note= "PCR product differs from gene sequence
 FT of AAF30046 at this base"
 XX
 XX WO200100838-A1.
 XX
 XX 04-JAN-2001.
 PD
 XX 23-JUN-2000; 2000WO-EP05854.
 PF
 XX 25-JUN-1999; 99GB-0015031.
 PR
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA Thonard J;
 PI
 XX WPI; 2001-112459/12.
 DR P-PSDB; AAB20107.
 XX
 XX Novel BASB110 polypeptides of Moraxella catarrhalis, useful as a
 PT vaccine for treating Moraxella catarrhalis infections -

XX Claim 10; Page 83; 88pp; English.

XX
 CC The present sequence is that of DNA encoding the BASB110 protein
 CC (see AAB20107) of Moraxella catarrhalis Mc2931 (ATCC 43617), a
 CC causative agent of otitis media in children and pneumonia in adults.
 CC The DNA was obtained by PCR amplification of Mc2931 plasmid DNA
 CC (see also AAF30048-49). The sequence differs from that of the
 CC BASB110 gene coding region sequence (see AAF30046) by having G at
 CC position 310 rather than A. This would cause an amino acid change
 CC from Gly-104 to Ser in the translated sequence. The invention
 CC provides BASB110 polypeptides, and polynucleotides encoding them,
 CC as well as expression vectors, host cells and methods for producing
 CC BASB110 polypeptides using recombinant methods. Also claimed is a
 CC vaccine composition comprising a BASB110 polypeptide, an immunogenic
 CC fragment of a BASB110 polypeptide, or a polypeptide having at least
 CC 85% amino acid sequence identity to BASB110, or comprising a
 CC polynucleotide encoding such a polypeptide. A claimed method of
 CC diagnosing a Moraxella infection involves identifying a BASB110
 CC polypeptide or antibody. A claimed therapeutic composition useful
 CC in treating humans with M. catarrhalis infection comprises at least
 CC 1 antibody directed against a BASB110 polypeptide. BASB110
 CC polynucleotides also have utility in diagnosis of the stage and
 CC type of infection, and also for therapeutic or prophylactic
 CC purposes, in particular genetic immunisation.

XX Sequence 966 BP; 263 A; 214 C; 242 G; 247 T; 0 other;

Alignment Scores:

Pred. No.: 3,71e-116 Length: 966
 Score: 1666.00 Matches: 321
 Percent Similarity: 99.69% Conservative: 0
 Best Local Similarity: 99.69% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 22 Gaps: 0

US-10-018-706-2 (1-322) x AAF30047 (1-966)

QY 1 MetThraThrlleAlleasenserglAaangllyserProilleysargleuGlyleu 20
 DB 1 ATGACTGTCAGATGACATGATTCACAAATCCCAACCAACCAACCAACCAACCAACCA 60
 QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerIysProThrTyrAsn 40
 DB 61 ATTTTGTGTGATCACCACTTGCAATTTTGCGAGATGTCCAGTAAGCCCAACCTTAAT 120
 QY 41 SerThrseryIserGlySerHisArgThrSeriYserGlyIleuAlaIleGlySer 60
 DB 121 AGTACCTCAGGTCGGGAGTCATCGTCAAGTTCAGGTCGATGTTGGCAATAGGTTCA 180
 QY 61 GlnValIleThrsPserGlnGlyValProAsnArgTyrGlnValIlyGlnGlyAspThr 80
 DB 181 CAGGTTATTCACGACATCGAGGCGGTACCAATCGCATACGAGGAAGGAGGCGCATACT 240
 QY 81 ValSerIysIleAlaGlnArgTyrGlyLeuAsnTrpArgIuIleGlyHisIleAsnAsn 100
 DB 241 GTAGTAAGATTGCTCAGCGTTATGATTAATGGCGGTGAGATTGGACACATTAAATAT 300
 QY 101 LeuAsnSerSerTyrThrIleIleTyrThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeu 120
 DB 301 CTAAATAGCGGTTATACGATTATACAGTCATAGCTGATCTTAATGTCAGGTGATCTC 360
 QY 121 LysValArgGlnArgSerIleSerSerGlyValAlaSerThrAlaHisThrProSerProVal 140
 DB 361 AAGGTGCGTACGACATACAGCTCTGTGTGTGATACAGCTCACACACCTTGCCCTGTG 420
 QY 141 AlaValAlnSerSerArgProProValGlnGlnHisProAlaValGlnIlyProThrPro 160
 DB 421 GCGGTTCATGTCAGACACACACAGTACACAGATCTCCCGTACAAAAAACCCACGCCA 480
 QY 161 ProValValValValIlyIlyIlyProThrProThrProProValValGlnGlnProAlaPro 180
 DB 481 CCTGT 540

QY 181 ValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValAlaMetGlnPhe 200
 DB 541 GTTGCCCCACCAAGTACAGAACACCACTTGGCCAGGGTAGCTCAGGGGTGTGCAATTT 600
 QY 201 ArgTyrProValGlyValAlaThrAsnProValValAlaArgPheGlyThrAlaThrValAla 220
 DB 601 CGCTATCTGTTGGTGGCGACCAATCCAGTGGTTGCACCGCTTGTGTACGGCGACAGTGCC 660
 QY 221 GlySerThrValThrSerSerGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsn 240
 DB 661 GGCTCAAGCTTTACAGATATGCAATGTGTTTCTTGACAGAGATGGGCAATTAACTAAC 720
 QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyValAspThrIleVal 260
 DB 721 GCGATGATGACAGGACAGTCACTTCAAGCTGATCAGATATGACAGGGCGAGTATTGTG 780
 QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValIly 280
 DB 781 ATTCAGCATACCAATGAGATTGTTCAGACTATATTCATATTACGACGCTCAGATTAA 840
 QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGly 300
 DB 841 ACAGGGGATACCGTGTGTCACCGTCAAGGTATTCAGAGCATGAAAAATCAGCCAAACCGGT 900
 QY 301 AlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValIlyValAspProLeuThrVal 320
 DB 901 GCGGCACTATTGGAATTGTGAATTTCTAGAAATGCGGTGTATGTTGATTCATTCACAGTA 960
 QY 321 LeuLys 322
 DB 961 CTTAA 966

RESULT 4
 AAT42063
 ID AAT42063 standard; DNA; 1830121 BP.
 AC AAT42063;
 NC 14-SEP-1999 (first entry)
 DT
 DE Haemophilus influenzae complete genome sequence.
 KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
 KW expression modulating fragment; regulation; gene expression; vector;
 KW organism; open reading frame; ORF; ds.
 OS Haemophilus influenzae.
 PN WO9633276-A1.
 PD 24-OCT-1996.
 PD 22-APR-1996; 96WO-US05320.
 PF 07-JUN-1995; 95US-0487429.
 PR 21-APR-1995; 95US-0426787.
 PR 07-JUN-1995; 95US-0476102.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (UYXO) UNIV JOHNS HOPKINS.
 PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
 DR WPI; 1996-485782/48.
 XX Haemophilus influenzae Rd genome recorded on computer readable
 PT medium - useful for identifying commercially important nucleic acid
 PT fragments by homology searching
 XX Claim 1; Page 77.2-77.1091; 1291pp; English.
 CC This sequence represents the complete genome sequence of the bacterium

Oy	153	ProAlaValGlnLysProThrProProValValValLysLysProThrProThrPro	172
Db	739	GCTGCGGTGCAAAACCCCTGTG-----AAACCGCGCGCGCAACCG	777
Oy	173	ProValValGlnGlnProAlaProValAlaPro-----	183
Db	778	CCTGGCAGTCCGCGCGCAACCTCCGCGCGCGGGAATAAAGCGGTTCCGCG	837
Oy	184	ProValThrGlnAlaProPheAla-----ThrGlySerSerGlyValMetGln	199
Db	838	CCCGCGCCGCAATCTCTCGCGCTTCGCGCTCCGCGCAACCGCTTCGCTCGCGCATTTGTT	897
Oy	200	PheArgTyrProValGlyAlaThrAsnProValValAlaGArgPheGlyThrAlaThrVal	219
Db	898	TGGCAGCTCCG-----ACGCAAGGTAAAGTGTTGCCGATTTCCGCGC-----	942
Oy	220	AlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspPheIle	239
Db	943	-----ACACAACAAGGTGTCGATTTATGCAGGAATAAGCGGACAGCCGCT	987
Oy	240	AsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer---	258
Db	988	TTGGGGGGGGCTGACGGCAAAAGTGTATATGACAGGTTCCGTTTACGGGATACCGCAAT	1047
Oy	259	---IleValIleGlnIleThrAsnGlyPheValSerSerTyrIleHisIleLysAspAla	277
Db	1048	TTGGTCATCATTCACGATTAATTTCTTCTTCGACCGCATACGGGACACCAACCAAAATTG	1107
Oy	278	GlnValLysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGln	297
Db	1108	CTGGTCGGCGAAGGCCACAGAGTCAAACGCGGCGAGCAGTCCGTTTATGGGCAATAC	1167
Oy	298	ProSerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspPro	317
Db	1168	GAGGCTTCTTAAGACGACGACCTTCATTTCGAGTGTGCGGCAAAACGCGTTAATCTCG	1227
RESULT 7			
AAZ53073			
ID	AAZ53073	standard; DNA; 1014 BP.	
XX	AAZ53073;		
AC			
XX			
DT	21-MAR-2000	(first entry)	
XX			
DE	Neisseria gonorrhoeae ORF 025 partial DNA sequence SEQ ID NO:117.		
XX			
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine; antigenic; diagnosis; immunogenic; infection; meningitis; septicemia; antibacterial; gene therapy; ds.		
XX			
OS	Neisseria gonorrhoeae.		
XX			
PN	W09957280-A2.		
PD			
XX	11-NOV-1999.		
PF			
XX	30-APR-1999;	99WO-US09346.	
XX			
PR	01-MAY-1998;	98US-0083758.	
PR	31-JUL-1998;	98US-0094869.	
PR	02-SEP-1998;	98US-0098994.	
PR	02-SEP-1998;	98US-0099062.	
PR	09-OCT-1998;	98US-0103749.	
PR	09-OCT-1998;	98US-0103794.	
PR	09-OCT-1998;	98US-0103796.	
PR	25-FEB-1999;	99US-0121528.	
XX			
PA	(CHIR) CHIRON CORP.		
XX	(GENO-) INST GENOMIC RES.		
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M, Petersen J, Piza M, Rappuoli R, Ratti G, Scarlato E, Scarselli M;		

[illegible]

```

Db      778  GTTATATCCCGTTCAGTTCAGGATACGGAATACTTGCATCATCCAGCAATTC
Qy      266  GlyPheValSerSerTyrIleHisIleAspAlaGlnValIysThrGlyAspThrVal
Db      838  TCTTCTCTGACCGCGTACGGGCAACCAAAATTGCTGTCGGCAGAGTCAGCAGTC
Qy      286  ArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheGlu
Db      898  AAACGGCTGACGAGCTGCTTCATGCGTAAATACCGATCTTCAGAACGACGCTTCAT
Qy      306  PheArgLysSerArgAsnGlyValIysValAspPro
Db      958  TTCGAGGTGCGTCAAAACGGCAACCGCTTACCCG

```

RESULT 8

AAZ53074 standard; DNA; 1056 BP.

AAZ53074;

21-MAR-2000 (first entry)

Neisseria meningitidis ORF 025 partial DNA sequence SEQ ID NO:119.

Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
antibacterial; gene therapy; ds.

Neisseria meningitidis.

WO957280-A2.

11-NOV-1999.

30-APR-1999; 99WO-US09346.

01-MAY-1998; 98US-0083758.

31-JUL-1998; 98US-0094869.

02-SEP-1998; 98US-0098994.

09-OCT-1998; 98US-0103749.

09-OCT-1998; 98US-0103794.

09-OCT-1998; 98US-0103796.

25-FEB-1999; 99US-0121528.

(CHIR) CHIRON CORP.

(GENO-) INST GENOMIC RES.

Fraser C, Galeotti C, Grandi G, Hickey E, Masiagnani V, Mora M;
Petersen J, Pizsa M, Rappoli R, Ratti G, Scalato E, Scarselli M;
Tetelini H, Venter JC;

WPI: 2000-062150/05.

P-PSDB; AA174312.

Novel Neisserial polypeptides predicted to be useful antigens for
vaccines and diagnostics

Claim 7, Page 206-; 1453p; English.

AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
PCR primers used in the exemplification of the present invention. The
polypeptides, the polynucleotides, antibodies and compositions of
the invention can be used as vaccines, as diagnostic reagents, and as
immunogenic compositions. The polypeptides can be used in the
manufacture of medicaments for treating or preventing infection due to
Neisseria bacteria (e.g. meningitis and septicemia), to detect the
presence of Neisseria bacteria, or to raise antibodies. They may also
be used to screen for agonists or antagonists, which may themselves
have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.

Sequence 1056 BP; 236 A; 322 C; 288 G; 210 T; 0 other;

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5.38e-06	198.00	41.51%	27.17%	11.86%	1056	72	38	95	60	12

US-10-018-706-2 (1-322) x AAZ53074 (1-1056)

```

Qy      73  TyrGlnValIysGlnGlyAspThrValSerLysIleAlaArgTyrGlyLeuAsnTrp
Db      361  CATACTATTGTGCGCGGACCGGTGTACCAATTCCTCAAGCTAC
Qy      93  ArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp
Db      409  -----CATACTCTCAAGACAT
Qy      113  LeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSerGly
Db      427  TTCGTCGTCGACGACGCG-----ATGACCGCAATATGTGACATCGTCGATT
Qy      131  -----ValAsnThrAlaHisThrProSerPro--ValAlaValGlnSerSerArgPro
Db      478  GTTAAAGTCAAAACCGGACGATATGCGCACCAAAACCGACCGGTGAAAGCAGGCC
Qy      148  ProValGlnGlnHisProAlaValGlnLysProThrProProValValValLysLys
Db      538  GCGGTACCG-----GCTGCGGTGCAAAACCCCTGTG-----AAA
Qy      168  ProThrProThrProProValValGlnGlnProAlaProValAlaPro
Db      571  CCGCGCGCGACACCGCTGTGACGTGCGCCGCCCAACTGCGCGCGCGCGGAAAT
Qy      184  -----ProValThrGluAlaProPheAla-----ThrGlySer
Db      631  AAACCGGTCCCGCGCGCGCGCGCGCAATCTCGCGCTTCGCGCGCGCGCGTTCG
Qy      195  SerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValValArgArgPhe
Db      691  GTCGCGCGCATGTGTTGGCAGCGTCCG-----ACGCAAGTAAAGTGTGCGCATTC
Qy      215  GlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArg
Db      745  GCGCGC-----AACCAACAGCGTGTGATTTGCCGTAT
Qy      235  AspGlyAspLeuIleAsnLysSerAsnAlaGlyThrValIleGluAlaAspHisMet
Db      781  GCGGACAGCCGCTTTTGGCGCGCGCGCGCAAGTGTATTATTCGCGTTCAGGTTG
Qy      255  AspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheValSerSerTyrIle
Db      841  AGCGATACGGAACATGTCATCATTCACCAATTTCTTCTTCTGACCCCATACGGG
Qy      273  HisIleLysAspAlaGlnValIysThrGlyAspThrValArgThrGlyGlnArgIleAla
Db      901  CACAAACCAAAATTTGCTGTCGCGCGGCGGCGAGGTCAACGCGGTACGACGAGTCT
Qy      293  SerMetLysAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAsnGly
Db      961  TTGATGGGCAAAATACCGATCTTCAGAACGACGCTTCATTCAGATGCTCAAAACGGC
Qy      313  ValTyrValAspPro
Db      1021  AAACCGTTTACCCG 1035

```

RESULT 9
AA81766/c
ID AA81766 standard; DNA; 6100 BP.

XX AAA81766;
 AC 04-DEC-2000 (first entry)
 DT
 XX N. meningitidis partial DNA sequence gnm_313 SEQ ID NO:313.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KM Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 PN MO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US23573.
 XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
 PI Masiarant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,
 PI Rapuoli R, Pizza M;
 XX WPI, 2000-318079/27.
 DR
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 PS Claim 7, Page 1588-1590; 1760pp; English.
 XX
 XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX
 SQ Sequence 6100 BP; 1552 A; 1757 C; 1455 G; 1336 T; 0 other;

Alignment Scores:

Pred. No.: 4.47e-05 Length: 6100
 Score: 198.00 Matches: 72
 Percent Similarity: 41.51% Conservative: 38
 Best Local Similarity: 27.17% Mismatches: 95
 Query Match: 11.86% Indels: 60
 DB: 21 Gaps: 12

US-10-018-706-2 (1-322) x AAA81766 (1-6100)

QY 73 TyrGlnValIysGlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrp 92
 Db 5739 CATACTATTTGGTGGCGGCGGACGAGTGTACACATTTCCTCAAGCGTAC----- 5652
 QY 93 ArgGlnIleGlyHisIleAsnLeuAsnSerSerTyrThrIleTyrGlnTrp 112
 Db 5691 -----CATATCTCTCAAGACGAT----- 5674
 QY 113 LeuThrLeuTrpSerGlyAspLeuLysValArgLysSerIleSerSerGly----- 130
 Db 5673 TTCGTCGCGTGAACGCGC-----ATGACCGACAATATGTGACCATCGTCAGATT 5623
 QY 131 -----ValAsnThrIleAsnIleThrProSerPro-----ValAlaValGlnSerSerArgPro 147
 Db 5622 GTTAAAGTCAACCGGACGATATGCGCACCGCAACCGACCGCGTGAAGAGGCGCC 5563
 QY 148 ProValGlnGlnHisProAlaValGlnLysProThrProProValAlaValLysLys 167
 Db 5562 GCGGTACCG-----GCTGCCGTGCACAAACCGCTGTG-----AAA 5530
 QY 168 ProThrProThrProProValAlaGlnGlnProAlaProValAlaPro----- 183
 Db 5529 CCGCGCGCGACACGCGCTGTGCACTCCGCGCGCACCTCGCGCGCGGCAAAAT 5470
 QY 184 -----ProValThrGluAlaProPheAla-----ThrGlySer 194
 Db 5469 AAAGCGTTCGCGCGCGCGCGCGCGCAATCTCTCGCTTCGCGCGCGCGCGCTTCG 5410
 QY 195 SerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValAlaArgPhe 214
 Db 5409 GTGCGCGCGCATTTGTGGCAGCGTCCG-----ACGCAAGTAAAGGTGGCGCATTTTC 5356
 QY 215 GlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArg 234
 Db 5355 GCGCGCG-----AACACAAAGGCTGCGATATTCGCGGTAT 5320
 QY 235 AspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMet 254
 Db 5319 GCGGACACGCGCGTGTGGCGCGCGCTGACGCGCAAAAGTGTATATTCGCGGTTCG 5260
 QY 255 AspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheValSerSerTyrIle 272
 Db 5259 AGGGAGTACGGAACATTGTCATCATCATCATCATCATCATCATCATCATCATCATCAT 5200
 QY 273 HisIleLysAspAlaGlnValIleThrGlyAspThrValArgThrGlyGlnArgIleAla 292
 Db 5199 CACAAACCAAAATTTGCTGTGCGCGGAGGCGGACAGGTCAACGCGGTCAAGGTTGCT 5140
 QY 293 SerMetLysAsnGlnProSerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsnGly 312
 Db 5139 TTGATGGGCAATTCGATGCTTCCAGACGACGCTCATTTCCAGAGTGCCTAAACGCGC 5080
 QY 313 ValTyrValAspPro 317
 Db 5079 AAACCGTTAACCGC-5065
 RESULT 10
 AAF21610/c
 ID AAF21610 standard; DNA; 349980 BP.
 XX
 AC AAF21610;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KM ds.
 XX
 OS Neisseria meningitidis.

Percent Similarity: 43.80%
 Best Local Similarity: 23.55%
 Query Match: 9.88%
 DB: 14
 Gaps: 6

US-10-018-706-2 (1-322) x ABQ51084 (1-1885)

QY 73 TyGlnValIysGlnGlyAspThrValSerIleAlaGlnArgIleuAsnTrp 92
 1229 TACAAGATGACCAAGGAGATGATGTTCTTATGTTGTTATTCAGGATGATATA 1288
 QY 93 ArgGluIleGlyHisIleAsnAsnLeuAsnSerSerIleIleIleIleIleIle 112
 1289 AAGAAATGGCCACATTAATATATGTCGAGCCATATCATCTGAGATGGAAGAATA 1348
 QY 113 LeuThrIleuTrpSerGly-----AspLeuIleValAlaGlnArgSerIleSerSer 129
 1349 TTGAAATGTCAAATTAATATTCGATGACATATGATGACACACACAAATTAATGAA 1408
 QY 130 GlyValAsnThrAlaHisIleThrProSerProValAlaValGlnSerSerArgProProVal 149
 1409 TCAGAGGTGACAAATATACGTCAATGAG--ACATGATGATGTTAATTAACCAAT 1465
 QY 150 GlnGlnHisProAlaValGlnIleProThrProProValValValValIleValProThr 169
 1466 GAACAAATGAACCCGTTGCTACACCAACATTCACCAATGCCAATCAATTAAT 1519
 QY 170 ProThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaPro 189
 1520 ---ACACCTCAGCCACCTCAATATATGCTTGATTTGGCCA----- 1558
 QY 190 PheAlaThrGlySerSerGlyValMetGlnPheArgIleProValGlyAlaThrAsnPro 209
 1559 -----ACAAATGGAATA 1570
 QY 210 ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
 1571 ATTAATTCAGATATTTCCAGGTGATGAGAGC-----AATTAAGGTATTT 1615
 QY 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr--ValIleG 249
 1616 GATATTAAGTCTTATGTCGATGACCAAGCTGTAATGCAAGCTGATGACCAAGCTTAT 1675
 QY 249 IlnAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisIleThrAsnGlyP 267
 1676 ATGCCGAGACGCTTACGTGATGATGTAATTAATTAATTAATTAATTAATTAATTAAT 1735
 QY 267 heValSerSerIleIleHisIleIleValAspAlaGlnValIleValIleValIleValIle 287
 1736 ATTAAAGTCTTATGACATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1795
 QY 287 hrgIleGlnArgIleAlaSerMetIleValAsnGlnProSerGlyValAlaLeuPheGluPhe 307
 1796 CGGGTCAACAAATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1855
 QY 307 rg 307
 1856 AA 1857
 DB
 RESULT 14
 ABQ67196
 ID ABQ67196 standard; DNA; 684707 BP.
 XX AC ABQ67196;
 XX DT 29-AUG-2002 (first entry)
 XX DE Listeria innocua contig DNA sequence #9.
 XX KM Antibacterial; Listeria; food contamination; mutational analysis;
 XX OS Listeria innocua.

XX W0200228891-A2.
 FN 11-APR-2002.
 XX PD 04-OCT-2001; 2001WO-FR03061.
 XX PF 04-OCT-2000; 2000FR-0012697.
 XX PR 04-OCT-2000; 2000FR-0012697.
 XX PA (INSP) INST PASTEUR.
 XX PA (CNRS) CNRS CENT NAT RECH SCI.
 XX PI Kunst F, Glaser P;
 XX DR WPI; 2002-332479/37.
 XX PT New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators
 XX PS Claim 5; SEQ ID 9; 180bp; French.
 XX CC The present invention relates to nucleic acid sequences
 CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
 CC and primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in
 CC anti-Listeria vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 684707 BP; 213423 A; 126375 C; 133438 G; 211468 T; 3 other;
 XX
 Alignment Scores:
 Pred. No.: 12.2 Length: 684707
 Score: 158.50 Matches: 77
 Percent Similarity: 36.09% Conservative: 41
 Best Local Similarity: 23.55% Mismatches: 127
 Query Match: 9.49% Indels: 82
 DB: 24 Gaps: 13
 US-10-018-706-2 (1-322) x ABQ67196 (1-684707)
 QY 27 ThrCysIleLeuAlaGlyCysAlaSerIleProThrTyraSerThrSerGlySerGly 46
 615017 AGTTGTATC---CGCAATGTTTAAGCGACCAAAATATATGCAATCAAAAAGACGA 615073
 DB
 QY 47 SerHisArgThrSerGlySer-----GlyGly 55
 615074 TCAAAATTAACATAAAGTACAGAAATATGTTCAAAAAAATTAAGAACTTAAGT 615133
 DB
 QY 56 -LeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrglnVa 75
 615134 GCTAGCAGTTGAGCTAGACAGTA-----TCAAGAAACACACTGAAAGCCAAACAA 615184
 DB
 QY 75 IlyGlnGlnIleAspThrValSerIleAlaGlnArgIleValLeuAsnTrpArgIle 95
 615185 AGAAAGAAATGACTTATGATGCTGCTGCAATTAATAAAGACTTAAG----- 615233
 DB
 QY 95 eGlyHisIleAsnAsnLeuAsnSerSerIleIleIleIleIleIleIleIleIle 115
 615234 -----AAAGTGAACAAACACTTTTAATTAATGA----- 615263
 DB
 QY 115 uTrpSerGlyAspLeuIleValArgGlnArgSerIleSerSerGlyValAsnThrAlaHis 135
 615264 -----CAAGGTGCTTTTCAACAGCAACAAAGAAACTGCTTCAATATGCTGATGAAA 615319
 DB
 QY 135 sThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVa 155

```

Db      615320 AGCAAAACAGAGCTGCAATTAAAGCGGCTGAAGAAAAAGCATGCAAGAGCAGCGGC 615379
Oy      155 1GlnlyProThrProProValValValValylsProThrProThrProProValVal 175
Db      615380 TAAAAATGCCACT-----GTGCGACAC 615400
Oy      175 1GlnGlnProAlaProValAlaProProValThrGluAlaProPheAlaThrGlySerse 195
Db      615401 GCAGCAACCAAGCTCTGTAAACATCTTCACAGAGAACTGGCGCAGACAGACGTAACTTC 615460
Oy      195 rGlyValMetGlnPheArgTyProValAla----- 206
Db      615461 TGGTGGCGGCGAGTTTATCAAGCCAGCATCAGAAATGTTAACTGCGATTAGTACGCG 615520
Oy      207 -ThrasnProValValAlaArgArgPhe-----GlyThrAlaThrValAlaGlySe 222
Db      615521 TACTAACCCAGTTACTGGAATAATATCAATCTCAATAAGTCAAGATATTGCGGCGGAGG 615580
Oy      222 rThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSe 242
Db      615581 TACGATTACA-----GTATCCGCTGC 615601
Oy      242 rAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
Db      615602 GGCATCTGTAGAGTCTGATTTTTCAGGATTGGCGCACAGGTAGCGGCTTGGAGGTTA 615661
Oy      259 -----IleValIleGlnHisThrAsnGlyPheValSerSerTyIleHisIlely 275
Db      615662 TGGTACGTCGTGAATAATGATCATGCTGTAATGCTTCCAAACATTATATGCTCATATGCG 615721
Oy      275 sAspAlaGlnValLys-----ThrGlyAspThrValArgThrGlyGlnArgIleAlaSe 293
Db      615722 TCCCGGTAGCTGAAAGTAGTAACTGCTCAACAGATACCAAGGACCAACCATTTGAT 615781
Oy      293 rMet-----LysAsnGlnProSerGlyValAlaIleuPheGluPheArgIleSerArgAs 311
Db      615782 TATGGGATCCACCGGTCATCAACCGGACACACATCTT---CATTTGAGATTTCACAAAAA 615838
Oy      311 nGlyValTyThrValAlaAspPro 317
Db      615839 TGGTATTCACGTTGATCCA 615857

RESULT 15
ABA03041/C
ID      ABA03041 standard; DNA, 2944528 BP.
xx      ABA03041;
AC      05-FEB-2002 (first entry)
DT
XX      Listeria monocytogenes EGD-e genome sequence.
DE
XX      Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KM      vitamin B12; bacterial infection; disease; ds.
XX      Listeria monocytogenes.
OS      WO200177335-A2.
XX      18-OCT-2001.
XX      11-APR-2001; 2001WO-FR01118.
XX      11-APR-2000; 2000FR-0004629.
PR
XX      (INSP) INST PASTEUR.
PA
PI      Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Denoux P,
PI      Dussauget O, Chetoui F, Nedjari H, Glaeser P, Kunst F, Cossart P,
PI      Daniels J, Goebel W, Krief J, Kuhn M, Ng B, Vazquez-Boland J,
PI      Domínguez-Bernal G, Garrido-García P, Tiberet-Martínez A, Amend A,
PI      Chakraborty T, Dommann E, Hain T, Berche F, Charbit A, Durant L,
PI      Perez-Diaz J, Baquero F, García Del Portillo F, Gómez-López N;

```

```

PI      Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI      Rose M, Voss H;
DR      MPI: 2002-010914/01.
XX
XX      Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT      and prevention of Listeria and related bacterial infections, and
PT      related polypeptides
XX
XX      Claim 1; SEQ ID No 1; 192bp; French.
XX
XX      The present sequence is the genome sequence of Listeria monocytogenes
CC      EGD-e. This sequence and fragments of this sequence are useful for
CC      selecting probes and primers for detecting genes in L. monocytogenes and
CC      related organisms, and to study genetic polymorphisms and other genomes.
CC      Proteins (AB047297-AB050149) expressed from the present sequence are
CC      useful for raising specific antibodies, identification of L.
CC      monocytogenes and related organisms, and for biosynthesis and
CC      biodegradation, especially polysynthesis of Vitamin B12. This sequence and
CC      proteins encoded by it are also useful for selecting compounds that
CC      regulate gene expression and cell replication and modulate L.
CC      monocytogenes-related diseases. In addition, this sequence and proteins
CC      encoded by it are useful in pharmaceutical and vaccine compositions for
CC      the treatment or prevention of infections by L. monocytogenes and related
CC      organisms.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
SQ      Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Alignment Scores:
Pred. No.: 71.2 Length: 2944528
Score: 159.50 Matches: 66
Percent Similarity: 37.22% Conservative: 33
Best Local Similarity: 24.81% Mismatches: 100
Query Match: 9.49% Indels: 67
DB: Gaps: 11

US-10-018-706-2 (1-322) x ABA03041 (1-2944528)
Oy      76 LysGlnGlyAspThrValSerLysIleAlaGlyArgTyGlyLeuAsnTrpArgGluIle 95
Db      2581153 GAAAAAATGATTTAGTATGCGCTTTCAGCAACAAAAAGACTTAATCT----- 2581106
Oy      96 GlyHisIleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeuThrLeu 115
Db      2581105 -----AAGAGTGAACAAACCTTTTACCTAGTGA----- 2581076
Oy      116 TrpSerGlyAspLeuLysValArgGluArgSerIleSerSerGlyValAlaLeuThrAlaHis 135
Db      2581075 ---CAAGTGCCTTACTGATGTAAGAAAAAAGACTTCTTAATGTTGCTGGTGA AAAA 2581019
Oy      136 ThrProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaVal 155
Db      2581018 GCAAAACAGAGGTGCAATTAAGTCCGGAAGAAAAAGCATGCAAGAGTGGCGCA 2580959
Oy      156 GlnLysProThrProProValValValValLysLysProThrProThrProProValVal 175
Db      2580958 GCATCAAGCGCTAAGACTGGCGCTGTGTCAA----- 2580926
Oy      176 GlnGlnProAlaProValAlaProProValThrGluAlaProPheAlaThrGlySerSer 195
Db      2580925 ---CAACCAAGTTCGTCATCAATAATGAAGCTACAGAA-----ACAGTAAAGTTCC 2580881
Oy      196 GlyValMetGlnPheArgTyProValAla----- 206
Db      2580880 GGTGAGAGCCCAATTTTAAACAGCTTCTGATTATTAACTTCGATTTAGTGAACGA 2580821
Oy      207 ThrAsnProValValAlaArgArgPhe-----GlyThrAlaThrValAlaGlySer 222
Db      2580820 ACAATTCCTCCGTGCTGCGCAATTCGAATCCCATTAAGTCAAGATATTGCTGCTGCTGTG 2580761

```

QY 223 ThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSer 242
 DB 2580760 ACAGTTACT-----GTTTCAGCTGCA 2580740
 QY 243 AsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
 DB 2580739 GCATCCGAAACAGTGTATTTCTGGGTTGGTCATCAGGTAGTGGATTGGTGTTAT 2580680
 QY 259 -----IleValIleGlnHisThrAsnGlyPheValSerSerTyrlleHisIleLeu 275
 DB 2580679 GCGTATGTAGTAAAAATTGACCAAGCGGTTCCAGACACTATATGACATATGGCT 2580620
 QY 276 AspAlaGlnValIle-----ThrGlyAspThrValArgThrGlyGlnArgIleAlaSer 293
 DB 2580619 GCAAGCAGATTGAAAGTATACCGGTCAACAGATTTCACAAAGCCCAACCAATTGGTATC 2580560
 QY 294 Met-----LysAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAsn 311
 DB 2580559 ATGGGATCAACAGGTCAATCGACAGCAACATCTA---CATTTGAATACATAAAAAC 2580503
 QY 312 GlyValTyArgValAspPro 317
 DB 2580502 GGTATTCAGTCAATCCA 2580485
 RESULT 16
 ABO69245/c
 ID ABO69245 standard; DNA; 3011208 BP.
 XX
 AC ABO69245;
 XX
 DT 29-AUG-2002 (first entry)
 XX
 DE Listeria innocua DNA sequence #684.
 XX
 KM Antibacterial; Listeria; food contamination; mutational analysis;
 KM infection; ds.
 XX
 OS Listeria innocua.
 XX
 OS WO200228891-A2.
 XX
 PN 11-APR-2002.
 XX
 PD 04-OCT-2001; 2001WO-FR03061.
 XX
 PF 04-OCT-2000; 2000FR-0012697.
 XX
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunst F, Glaser P;
 XX
 DR WPI; 2002-332479/37.
 XX
 PT New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators -
 XX
 PS Claim 5; SEQ ID 2058; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (ABO67188-ABO71212) from Listeria sp. The sequences are useful as probes
 CC and primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in
 CC anti-Listeria vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;
 Alignment Scores:
 Pred. No.: 73.1 Length: 3011208
 Score: 158.50 Matches: 77
 Percent Similarity: 36.09% Conservative: 41
 Best Local Similarity: 23.55% Mismatches: 127
 Query Match: 9.49% Indels: 82
 DB: 24 Gaps: 13
 US-10-018-706-2 (1-322) x ABO69245 (1-3011208)
 QY 27 ThrCysIleLeuAlaGlyCysAlaSerLysProThrTyraSerSerThrSerGlySerGly 46
 DB 2665796 AGTTGTATC---CGCAATGTTTAAAGGAGCAAAATATATGCAAGATCAAAAAGCA 2665740
 QY 47 SerHisArgThrSerGlySer-----GlyGly 55
 DB 2665739 TCAAAATTAACCTAAAGTAGCAGAAATATGTCAGAAAAAAATTAAGAACTTAAGT 2665680
 QY 56 -LeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrglnVal 75
 DB 2665679 GCTAGCAGTTGAGCTTAGAAGTA-----TCAGAAACCAACATGGAAAGCCAAACAA 2665629
 QY 75 LlyeGlnGlyAspThrValSerLysIleAlaGlnArgTyrglyLeuAsnTrpArgGluI 95
 DB 2665628 AGAAAGAAATGACTTATGATGCTCTTGCAAAATAAAGACTTAAG----- 2665580
 QY 95 eGlyHisIleAsnAsnLeuAsnSerSerTyrlleTyrlleTyrlleGlnTrpLeuThrLe 115
 DB 2665579 -----AAAGTGAAACAAACACTTTTAATGAA----- 2665550
 QY 115 uTrpSerGlyAspLeuLysValArgGluArgSerLysSerSerGlyValAsnThrAlaHi 135
 DB 2665549 ----CAAGTGCTCTTTCACAGCAAGAAAGAAACCTTCTTAATATGCTGTGTA 2665494
 QY 135 sThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVa 155
 DB 2665493 AGCAAAACAAAGAACCTCAATTAAAGCGCTGAAGAAACGCTGCAGAAAGCAGCCG 2665434
 QY 155 lGlnLysProThrProProValValValLysLysProThrProThrProProValVa 175
 DB 2665433 TAAAAATGCCACT-----GTCCGACG 2665413
 QY 175 lGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySerSe 195
 DB 2665412 GCAGCAACCAAGCTCTGTATCATCTTCAGCAGAAATGGCGCAACTGACACAGTAAGTTC 2665353
 QY 195 rGlyValMetGlnPheArgTyrrProValGlyAla----- 206
 DB 2665352 TGGTGGCGGCGCACTTTATCAAGCCAGCATCAGAAATTTAACTCTGATTTAGTGACG 2665293
 QY 207 -ThrAsnProValValArgArgPhe-----GlyThrAlaThrValAlaGlySe 222
 DB 2665292 TACTTAACCCAGTTACTGGAATAATATGATCTCAATAAGGTCAGATATTTGGCGGCGAGG 2665233
 QY 222 rThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSe 242
 DB 2665232 TACGATTACA-----GTATCCGCTG 2665212
 QY 242 rAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
 DB 2665211 GGCATCTGGTAGAGCTGATTTTTCAGATTGGCGCAACAGTACGGCTTTGAGAGTTA 2665152
 QY 259 -----IleValIleGlnHisThrAsnGlyPheValSerSerTyrlleHisIleLeu 275
 DB 2665151 TGGTATAGTGTGTAATAATATGATGATATGCTTCAAAACATTTATGCTCATATGCG 2665092
 QY 275 sAspAlaGlnValIle-----ThrGlyAspThrValArgThrGlyGlnArgIleAlaSe 293
 DB 2665091 TGCCGGTAGCTTGAAGTATTAATGCTCAACAGATATCAAGAGCAACCAATTTGGTAT 2665032
 QY 293 rMet-----LysAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAs 311

```
Db 265031 TATGGGATCCACCGGTCAATCAACCGCAACATCTT---CATTTGAGATTCACAAA 2664975
QY 311 ngiVal1YrValAspPro 317
Db 2664974 TGGTATTCAGTGTGATCCA 2664956

RESULT 17
ID ABA90521/c
XX ABA90521 standard; DNA; 2365589 BP.
AC ABA90521;
XX
DT 16-MAY-2002 (first entry)
DE Genomic sequence of Lactococcus lactis IL1403.
XX
KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
DR WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification or Lactococcus
XX
PS Lactis and related species -
XX
XX
XX Claim 1; SEQ ID 1; 2504bp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (AB53300-AB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO20017733 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at tcp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

Alignment Scores:
Pred. No.: 398 Length: 2365589
Score: 147.00 Matches: 77
Percent Similarity: 38.22% Conservative: 43
Best Local Similarity: 24.52% Mismatches: 112
Query Match: 8.80% Indels: 83
DB: 24 Gaps: 12

US-10-018-706-2 (1-322) x ABA90521 (1-2365589)
QY 3 ValThrIleAlaIleAsnSerGlnAsnGlnLysProIleLys----- 16
Db 269567 ATTAATGCTAAAGCTTGGCGAACAACGCGCACTTAACAGCTGCAACCGCACCTTAA 269508
QY 17 -----ArgIleuGlyIleuIlePheGlyValIleThrTrpGlyIleuAlaGly 32
Db 269507 CAGGAAATATGACAACTGACCTAATTAT-GGTGCTTCCGATCGAATCAATTATTTCT--- 269452
QY 33 CysAlaSerLysProThrTrpLysAsnSerThr-----SerGlySerGlySerHisArgThr 50
```

```
Db 269451 -----CAATATATATTTGCTGCTTTGACGGTCTTCTTCTGCTGACT 269407
QY 51 SerGlySerGlyIleuAlaIleGlySerGlnValIleThrAspSerGlnGlyValPro 70
Db 269406 TCTATATTCGGTGGT---TCAACAGCTTCAAAATACCAATATATATCAATATCAAGCTCA 269350
QY 71 AsnArgGlyTrpGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTrpGlyIleu 90
Db 269349 ACCACTTATACAGTATTAATCTGGGGAATACCTTTGGGAATTTGCAAAATATGGAATT 269290
QY 91 AsnTrpArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTrpTrpIleTrpThrGly 110
Db 269289 AGTGTGCTCAAAATTCAAAGCGCAACAAATCTTAAAGTACA---GTCACTTATATTGGG 269233
QY 111 GlnTrpLeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSerSerGly 130
Db 269232 CAAAGCTTGTATTTGACAACTTCAAGTCTTCTGCTAATACAAATAGTCAACTTCTTCA 269173
QY 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
Db 269172 GGAATTTCTGCCGGAACACTCAACCGCTACTACT----- 269140
QY 151 GlnHisProAlaValGlnLysProThrProProValValValLysLysProThrPro 170
Db 269139 -----TCGGTCACCTCT 269128
QY 171 ThrProProValValGlnGlnProAlaProValAlaProProValThr----- 186
Db 269127 GCCAAACGAGCTTACAGAGAGACATTAAGTTAATCTGTGATACGCTTGGGACATC 269068
QY 187 GluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTrpProValGlyAla 206
Db 269067 TCTGTCAATATTAAGACAGATTCCTCACTCAAGTGGATTCATTGAATCTTGAT 269008
QY 207 ThrAsnProValValArgTrpPheGlyThrAlaThrValAlaGlySerThrValThrSer 226
Db 269007 ACAATTTTCATGGACAAACCTTGATTTTTCACAAATCTCCGATCTTCTTCAAGTCTTCA 268948
QY 227 AsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr 246
Db 268947 ACACGT-----TCAAGCTCAGCCTCTTACG 268924
QY 247 ValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHisThrAsnGly 266
Db 268923 -----AGTTCAACTTCTTAACCTCTTCTGACGCTTCA----- 268894
QY 267 PheValSerSerTrpIleHisIleLysAspAlaGlnValLysThrGlyAspThrVal--- 285
Db 268893 -----AATACCTCTATCCAT-----AAGTTGTATTAAGAGATACGCTTTGG 268852
QY 286 -----ArgThrGlyGlnArgIleAlaSerMetLys 295
Db 268851 GGACTTTCACAAAATCTGTGTAGCCCAATTGCTTCAATTAAAG 268810

RESULT 18
ID AAH67706
XX AAH67706 standard; DNA; 615 BP.
AC AAH67706;
XX
DT 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 2741.
XX
KM Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
XX 20-JUN-2001.
```

XX 18-DEC-2000; 2000EP-0127688.
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI; 2001-376931/40.
 XX P-PSDB; AAG92487.
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 8; SEQ ID NO: 2741; 246bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 615 BP; 148 A; 229 C; 139 G; 99 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.0204 Length: 615
 Score: 146.50 Matches: 64
 Percent Similarity: 44.44% Conservative: 20
 Best Local Similarity: 33.86% Mismatches: 67
 Query Match: 8.77% Indels: 39
 DB: 22 Gaps: 10
 US-10-018-706-2 (1-322) x AAH67706 (1-615)
 QY 137 ProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
 Db 107 CCCGACCGCTGACG-----CCACCATGGGACCGACCATCG-----CGC 145
 QY 157 LysProThr-----ProProValAlaValValValValValValValValValValVal 172
 Db 146 AAACCGCATGCGAGCGCTTCTCCCGCGCTG-----CCGACACCGCTCCGACG 193
 QY 173 ProValAlaGln---GlnProAlaProValAlaProProValThnGlnAlaProPha 191
 Db 194 AAGCTGAGAGATGCGACACCGCGCTC-GCACACCGCGCGACCGGACCGCGCGCA 252
 QY 192 ThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAspProVal 211
 Db 253 AACGGGACCC-----TTCACTCTCAGGATTGGGA 279
 QY 212 ArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerArgMetTyrPhe 231
 Db 280 CCAAGTTGGGAAAC-----TTCCACCAACGCGATCGACATC 315
 QY 232 SerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla 251
 Db 316 GCAATCATCATCGGACCCCATCTACCGCTCATGCGCGACATGATGAGTCTGGC 375
 QY 252 HisAsnMetAsp---GlyAlaSerIleValIleGlnHisThrAsnGlyPheValSer 270

Db 376 CCAGCATCCGGCTATGACAGTGGATCCGATCCAGACGACGACGATCCATCCATC 435
 QY 271 TyrIleHisIleLeuAspAlaGlnValLysThrValLysThrValAlaGlyThrGlyAla 290
 Db 436 TACGACACATGAAATCTCTACGCTCCGCGCAACCGCTCGGACGACGACGAGAA 495
 QY 291 IleAlaSerMetLysAsnGlnPro---SerGlyAlaAlaLeuPheGluPheArgIleSer 309
 Db 496 ATCCGAGATTTGGGACACCAAGATTCTCCACCGCGCTCCACCTCCACTTCGAATTCAC 555
 QY 310 ArgAsnGlyValTyr---ValAspPro 317
 Db 556 CCAGACGCGCTCACCCCGATCGACCA 582
 RESULT 19
 AAH68532
 ID AAH68532 standard; DNA; 349980 BP.
 XX
 AC AAH68532;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 7067.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI; 2001-376931/40.
 XX
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Disclosure; SEQ ID NO: 7067; 246bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 349980 BP; 80900 A; 98397 C; 92139 G; 78544 T; 0 other;
 Alignment Scores:
 Pred. No.: 43.3 Length: 349980
 Score: 146.50 Matches: 64

Percent Similarity: 44.44%
 Best Local Similarity: 33.86%
 Query Match: 8.77%
 DB: 22
 Gaps: 10

US-10-018-706-2 (1-322) x AAF59421 (1-349980)

QY 137 ProserProValAlaValGlnSerSerProValGlnGlnHisProAlaValGln 156
 DB 249656 CCGACACGCTACCG-----CCACACGACGACCGACGCTCG-----CGC 249694
 QY 157 LysProThr-----ProProValAlaValValValValValProThrProThrPro 172
 DB 249695 AAGCCGACGACGACGACGCTTCTCCGCGCTG-----CCGACACGCTCGCAG 249742
 QY 173 ProValValGln-----GlnProAlaProValAlaProProValThrGlnAlaProPheAla 191
 DB 249743 AAGCTGACGAGTGCACACCGCCGTC-GCACACGCGCGACCGCTAGCGCCGACGCA 249801
 QY 192 ThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValVal 211
 DB 249802 AACGGCAC-----TTCCACCTCAGGATTCGCA 249828
 QY 212 ArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPhe 231
 DB 249829 CCACGTTGGGGAAC-----TTCCACACGCGCATCGACATC 249864
 QY 232 SerGlyArgAspGlyAspLeuLeuLeuAsnAlaSerAsnAlaGlyThrValIleGlnAlaAsp 251
 DB 249865 GCACACGACATCGGACACCCCAATCTACGCGCTCATGCGCGCATCTCATCTCACTCTCGC 249924
 QY 252 HisAsnMetAsp-----GlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSer 270
 DB 249925 CCAGCATCCGGCTATGACAGTGCATCGCATCGACGACGACGACGATCATCTCATCTC 249984
 QY 271 TyrIleHisIleValAspAlaGlnValIleThrGlyAspThrValAlaThrGlyGlnArg 290
 DB 249985 TACGGACACATGAAATCTCTTACCTCTCCGTCGCAACCGCTCGACGACGACGACGAA 250044
 QY 291 IleAsnSerMetLysAsnGlnPro-----SerGlyAlaAlaLeuPheGlnPheArgIleSer 309
 DB 250045 ATCCACAGAAATGGGACGACGACGATTCCTCCACCGCTCCACCTTCGATTCGATTCAC 250104
 QY 310 ArgAsnGlyValTyr-----ValAspPro 317
 DB 250105 CCAGACGCGCTCACCCCGATCGACCCA 250131

RESULT 20
 AAF59421/C
 ID AAF59421 standard; DNA; 949 BP.
 XX AAF59421;
 AC AAF59421;
 XX 02-MAY-2001 (first entry)
 DT 02-MAY-2001 (first entry)
 XX
 DE Actinobacillus actinomycetemcomitans clone nucleotide sequence #8.
 XX
 KM Actinobacillus actinomycetemcomitans; microbial; infection; vaccine;
 KM identification; localised juvenile periodontitis; antibacterial;
 KM antiinflammatory; de.
 XX
 OS Actinobacillus actinomycetemcomitans.
 XX
 MO200111081-A2.
 XX
 PD 15-FEB-2001.
 XX
 PF 04-AUG-2000; 2000WO-US21340.
 XX
 PR 06-AUG-1999; 99US-0147551.
 XX
 PA (IVIG-) IVIGENE CORP.
 XX

PI Proguiske-Fox A, Handfield M, Brady LJ, Hillman JD;
 DR WPI; 2001-202779/20.
 XX
 PT Identifying microbial polynucleotides, useful for vaccine design,
 PT diagnostic and antibiotic therapy, comprises isolating clones of a
 PT microbe's expression library, isolating clones of a microbe's
 PT protein produced during in vivo growth -
 PS Example 3; Page 63; 68pp: English.

CC The present invention describes a method (M1) for identifying a
 CC polynucleotide (I) of a microbe (M) that is expressed in vivo. (M1)
 CC comprises: (a) absorbing antibodies (Ab) against antigens that are
 CC expressed by (M) in vivo and in vitro with cells or cellular extracts
 CC of (M) that have been grown in vitro; (b) isolating unadsorbed Abs; and
 CC (c) probing an expression library of (M)'s DNA/RNA with unadsorbed Abs,
 CC where (I) that is expressed in vivo is identified. The method can be
 CC used for identifying antigens expressed during an actual microbial
 CC infection. The identified polynucleotides are useful for vaccine design,
 CC diagnostic and antibiotic therapy, in particular for the diagnosis and
 CC therapy of Actinobacillus actinomycetemcomitans infection, which is the
 CC etiologic agent for localised juvenile periodontitis. The present
 CC sequence represents an Actinobacillus actinomycetemcomitans clone
 CC nucleotide sequence, which is used in an example from the present
 CC invention.

SO Sequence 949 BP; 170 A; 236 C; 281 G; 261 T; 1 other;

Alignment Scores:
 Pred. No.: 0.0447 Length: 949
 Score: 145.00 Matches: 57
 Percent Similarity: 37.04% Conservative: 23
 Best Local Similarity: 26.39% Identical: 62
 Query Match: 8.68% Indels: 74
 DB: 22 Gaps: 9

US-10-018-706-2 (1-322) x AAF59421 (1-949)

QY 73 TyrGlnValIleGlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrp 92
 DB 537 TATACCGTACGCAAGGTGACACCAATGCTTATCGCTCATCTTCTGTTGGATGTG 478
 QY 93 ArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112
 DB 477 AAGAAATGGCGCGCGCTGAATATATGTCGCAACCTTACAGCTTAAGGTGCAACAAAC 418
 QY 113 LeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSerSerGlyValAsn 132
 DB 417 TTA-----AAGTGAGCCGCGCG----- 400
 QY 133 ThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152
 DB 399 -----ACAAAGCTTACGGTGCACAAAGAAAGCTGCACCCCAATGCA----- 358
 QY 153 ProAlaValGlnLysProThrProProValValValValValValProThrProThrPro 172
 DB 357 -----CCAGCGCTCACCCA 343
 QY 173 ProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThr 192
 DB 342 CCGGCTACGCAAGGTGTGACAC-----GTAACCTTACACCCCGCGCGCAAC 292
 QY 193 GlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValValArg 212
 DB 291 GGCACA-----CAATATGCTTCCAGGACGACGATTAACCGCGCGCTAAAGCC 244
 QY 213 ArgPheGlyThrAla-----ThrValAlaGlySerThr----- 223
 DB 243 GCGGTAGTACTGTGACCGGCAACCAATCAGCGATTAACCAAGCGGCGCACCGACCG 184
 QY 224 ValThrSerAsnGlyMetTrp-----PheSer 232

Db	183	GCAGATATCCAAAGTGGCATGGGCGTGGCCGACCAAGCGCAATGTATTCAGAGTTTCTCT	124
Qy	223	GIyATgApGlyI-----AspleuIlleasn	240
Db	123	AACGCCGACGGCGGTACCAAGGACATGCATAGTGTCTTCGACAGCAAGAGCTCAAT	64
Qy	241	AlaseAenAlaGIyThrValIlleGlnAlaSpHIsaMetAepGly	256
Db	63	GCCGCGCGCGGATCGAGTGGTTTACGCGGGTAAAGCCTTACGCGGT	16
RESULT 21			
ID	AAx20500/c		
AC	AAx20500 standard, DNA; 14063 BP.		
XX			
XX	AAx20500;		
XX			
DT	05-MAY-1999 (first entry)		
XX			
DE	Polynucleotide sequence from the genome of Treponema pallidum.		
XX			
KW	Treponema pallidum infection; syphilis; Borrelia infection; animal;		
KW	enzyme production; ds.		
OS	Treponema pallidum.		
PN	WO9859034-A2.		
XX			
PD	30-DEC-1998.		
XX			
PF	23-JUN-1998; 98WO-US13041.		
XX			
PR	24-JUN-1997; 97US-0050667.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Fraser CM;		
XX			
DR	WPI; 1999-081273/07.		
XX			
PT	New isolated Treponema pallidum nucleic acids - used to develop		
PT	products for the detection, diagnosis, characterisation, prevention		
PT	and therapy of T. pallidum infections, particularly syphilis		
XX			
PS	Claim 1; Page 178-186; 1150pp; English.		
XX			
CC	AAx20500-21243 represent polynucleotide sequences from the genome of		
CC	Treponema pallidum. The sequences can be used for detection,		
CC	diagnosis, characterisation, prevention and therapy for T. pallidum		
CC	infections, particularly syphilis. They can also be used for detecting		
CC	diseases related to Borrelia infections in animals, and for the		
CC	production of biosynthetic products such as enzymes.		
XX			
SQ	Sequence 14063 BP; 3225 A; 4086 C; 3507 G; 3225 T; 20 other;		
Alignment Scores:			
	Pred. No.:	2.75	Length: 14063
	Score:	140.00	Matches: 82
	Percent Similarity:	38.62%	Conservative: 52
	Best Local Similarity:	23.63%	Mismatches: 141
	Query Match:	8.38%	Indels: 73
	DB:	20	Gaps: 14
US-10-018-706-2 (1-322) x AAX20500 (1-14063)			
Qy	23	GIyValIlleThrCysAlleuAlaGlyCysAlaSerIyProThrTyrAsnSerThr	42
Db	8443	GGAACCATTTGACACCTGCGCTTCACGCTCACTGAGCA-----GGAACCAAG	8399
Qy	43	SerGIySerGIy-----SerHIsaArgThrSerGIySerGIy	54
Db	8398	TCAGGCGTCGCGGATCGTGTCTTCTGCTCCGTCACAGTATCGTCAGAAATGCGGA	8333
Qy	55	GIyLeuAlaIlleGIySerGIyValIlleThrAspSerGIyValIProAsnArgTyrGln	74

Dd	8338	GCCTGCACTGACTACACAGCTGA	----	CGA	8312				
Qy	75	VallysgInglYAspThrValSerIleAlaGlnArgIYLeuAsnTrpArgIu	94						
Dd	8311	ATGCGTGAAGGTGATGTGTGAGAAGATTGCCAGCGCTATGACATCAGCAGATGCA	8252						
Qy	95	IIeGIYhSIleAsnIleuAsnSerSerYrThrIleYrThrGInTrpLeuThr	114						
Dd	8251	ATCATTAAGTTGAATTAATATGCGCACAACCG-CCGCTCCAAAGTTGACAGCTACTAATA	8193						
Qy	115	LeuTrpSerGIYAsp-----LeuYsValArgGlnArgSerIleSerSerGIYVal	131						
Dd	8192	ATCCCTCAAGTGCACCGCATTTTATACCTGATAAAAAAGCGCATAGCTTTGCTCATATA	8133						
Qy	132	AsnThrAlaIleThrProSerProValAlaValGlnSerSerArgProProValGlnGln	151						
Dd	8132	GGCGGGCGGCAT-----CAGATCTCCTTAAGAACATATGTCCTCTTAAC	8088						
Qy	152	HisProAlaValGlnIYProTrhProProValVal---ValValIYsIYsPro-----	168						
Dd	8087	ACGCGCTCTCTTCAAGAGAGTCACTCTTCAAGTGGGTAAGCGTGGTTTCTCGCTCAT	8028						
Qy	169	-----ThrProThrPro-----	172						
Dd	8027	AATTCACTGACAGCGGAGTCTCGCGTCTTTCCGTTTCTTCTCGGAAACAGTGAAG	7968						
Qy	173	-----ProValValGln-----GlnProAlaProValAlaProProVal	185						
Dd	7967	GAATAATACGTGTTTGAACGACGATCAGCATGACGCTGA-----CGGTA	7920						
Qy	186	ThrGlnAlaProPheAlaIleThrGlySerSerGIYValMetGln-----	199						
Dd	7919	CTGTTTTTGGCCCGGTGACATTTAAGCGCAGCCGCTTGACGAATAATATGTGACTTG	7860						
Qy	200	PheArgTrpProValGlyAlaIleThrAsnProValValArgArgPheGly-----ThrAla	217						
Dd	7859	TTTGGTCTCTCTTCGAGTGGTACTATGTTTCTTCGCGTATGATGGAGCGTATGAT	7800						
Qy	218	ThrValAlaGlySerThrValIleThrSerAsnGlyMetTrpPheSerGIYArgAspGIYAsp	237						
Dd	7799	CCGTTTACCGGTGCAGAAAGTTTACAAATGTTTGATATGTGAGTCCGCGGGTACC	7740						
Qy	238	LeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsn---MetAspGIY	256						
Dd	7739	CCAGTATATTCGGCGCTCGAGGAGGATGTATGACGACGGTTGGATACAGTGCAGTGTATG	7680						
Qy	257	AlaSerIleValIleGlnHisThrAsnGlyPheValSerSerYrIleHisIleIYAsp	276						
Dd	7679	AATTACCTGATTTGGGGCACACGCGGGGTATACAGCCCTGATGGGCATCTCAACAG	7620						
Qy	277	AlaGlnValIYsThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMet---IYs	295						
Dd	7619	GTTGTTGTTTACAGAGTACCGCGGTACACACGCGCAAAAAATCGATTATTTGGGAAA	7560						
Qy	296	AsnGlnProSerGIYAlaIleAlaLeuPheGlnPheArgIleSerArgAsnGIYValTrpVal	315						
Dd	7559	ACAGAGCAGCAGCGGACCAACACCTGATTTTCAATCTATATAGATGGCTCGCATTA	7500						
Qy	316	AspProIleuThrValIleIYs	332						
Dd	7499	AACCTTACTCTCTTACTGCGT	7479						
RESULT 22									
AB067197/c									
ID	AB067197	standard; DNA; 1163020 BP.							
XX	AB067197;								
XX	AC								
XX	DT	29-AUG-2002 (first entry)							
DE	Listeria innocua contig DNA sequence #10.								
XX									

SQ Sequence 1233 BP; 468 A; 225 C; 241 G; 299 T; 0 other;

Alignment Scores:

Pred. No.:	0.41	Length:	1233
Score:	134.00	Matches:	62
Percent Similarity:	39.29%	Conservative:	37
Best Local Similarity:	24.60%	Mismatches:	111
Query Match:	8.02%	Indels:	42
DB:	24	Gaps:	10

US-10-018-706-2 (1-322) x AAL46626 (1-1233)

Qy	74	GlnValLysGlnIleAspThrValSerLysIleAlaGlnArgLysIleAsnThrParg	93
Db	562	CAAAAACAAACAAACAAACATTCGCAAAAAGCAGCAAGCAAGCATCAATTTTCG	615
Qy	94	GluIleGlyHisIleLeuAsnLeuAsnSerSerTyrrTrileTyrrThyGlnIlePhe	113
Db	616	-----CTGAATGACTCACTAATAAATTAGCCCTAGATCAAGATAAA-----	657
Qy	114	ThreLeuTrpSerGlyAspLeuLysValAlaGlnArgSerIleSerSerGlyValAsnThr	133
Db	658	-----TTGAATGACTATAAGCAACGACAAACGACACTTCCTGCAAGAAATCAACGA	708
Qy	134	AlaHisThrProSerProValAlaValAlaGlnSerSerArgProProVal-----GlnGlnHis	152
Db	709	GCGAACAACGACGACCGGCAACAGAAACCTGAAAGAGGCACTGCTTCCAAAGCGCAA	768
Qy	153	ProAlaValGlnLysProThrProProValValValLysLysProThrProThrPro	172
Db	769	AAAGCTGAAGAAAAAGAACAA-----	789
Qy	173	ProValValGlnGlnProAlaProAlaProAlaProAlaProPheAlaThr	192
Db	790	-----TCMAAACCTTATCAACCAACCTGACAAAGCCCAATTACTTAAT	834
Qy	193	GlySerSerGlyVal-----MetGlnPheArgTyrrProValGlyAlaThrAsn	208
Db	835	AGTACAAAGCGGTTTAGCGGGCGCAAAAAAACAAATATCTTACCAGTTTCGGTTCA--	891
Qy	209	ProValValArgAspPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGly	228
Db	892	----ATTTCGATCACTTTTGT-----TCATACCAAGACGGCAAGTACGTGGAAAGGT	942
Qy	229	MetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle	248
Db	943	ATGTGTAATTTGGTCATCAGCAGGCGACCGCTGTAAAGCAATTCGACGTGGACGGCTATT	1000
Qy	249	GlnAlaAspHis---AsnMetAspGlyAlaSerIleValIleGlnHisThrAsnGlyPhe	267
Db	1003	TTAGCGGAGATTAAATGATTATGTTATATGTTATATGTTATTAACACGGCGAAATCGAT	1060
Qy	268	ValSerSerTyrrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThr	287
Db	1063	TTTAGCTTTATATGGCTTCATCAACACGGGTATCAAGAAAGTGTGTCAGCTTGTTTCGCA	1120
Qy	288	GlyGlnArgIleAlaSerMetLysAsn-----GlnProSerGlyAlaAlaLeuPheGlu	305
Db	1123	GGGCAAGGTAATTCCTCAAGTACGAAATACAGGGGAAATATACAGTTCGGCGCTTAT---	117
Qy	306	PheArgIleSerArgAsnGlyValLysValAspPro	317
Db	1180	TTTGATTTAGCCGTAAAGCAACGCCAGTAATATCA	1215
RESULT 25			
AAL46624			
ID	AAL46624 standard, DNA: 1233 BP.		
AC	AAL46624;		
XX			
DT	05-ANG-2002 (first entry)		
XX			
DE	H influenzae BARS201 coding sequence #1.		

XX	BASB201; otitis media; pneumonia; sinusitis; nosocomial infection,
KM	auditive nerve damage; delayed speech learning; vaccine,
KM	antibacterial; auditory; antiinflammatory; gene; ds.
XX	
OS	Haemophilus influenzae.
XY	

FH	Key	Location/Qualifiers
FT	CDS	1.1233
FT		/*tag= a
FT		/product= "BASB201"
..		

PN WO200230967-A2

PD 18-APR-2002
YY

PF 05-OCT-2001; 2001WO-EP11561.
XX

PR 13-OCT-2000; 2000GB-0025169.
XXPA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
XX

XX. Томара 2,

DR WFL; 2002-426261/45
DR P-PSDB; AA017656.

AA PT	New isolated non-
----------	-------------------

as otitis media, delayed speech learning and inflammation of middle ear

PS Claim 13; Page 87; 90pp; English.

The present invention provides the protein and coding sequences of several versions of the BARS201 protein from non-typable *Haemophilus influenzae*. These can be used in the production of vaccines against *H. influenzae* infection, which can cause otitis media in infants and children, pneumonia in elderly, sinusitis, nosocomial infections, or invasive diseases, chronic otitis media with hearing loss, fluid accumulation in the middle ear, auditory nerve damage, delayed speech learning, infections of the upper respiratory tract and inflammation of the middle ear. The present sequence is a version of the BARS201 coding sequence of the invention.

SQ Sequence 1233 BP; 466 A; 224 C; 239 G; 304 T; 0 other;

Alignment Scores:	
Pred. No.:	0.688
Score:	131.00
Percent Similarity:	39.89%
Best Local Similarity:	22.6%
Query Match:	7.84%
	Indels:
	1233
	Matches:
	62
	Conservative:
	36
	Mismatches:
	112
	Indels:
	42

US-10-018-706-2 (1-322) x AAL46624 (1-1233)

[illegible]

Db 3538 ATGCCGAGACGCTTACGTGATATGTAATTAATTATTATTAACAATGACAGTT 3597
QY 267 hevalsergyllehis 273
3598 ATTTAGTGGCTTATGCAAT 3617
RESULT 27
AA06135
ID AA06135 standard; DNA; 1929 BP.
XX AA06135,
XX
XX 01-APR-1996 (first entry)
XX
XX Lysin coding sequence.
XX
XX Lysin; autolysis; culture; lactic acid bacteria; fermentation;
XX cheese, foodstuffs; induction; ds.
XX
XX Lactococcus lactis (Strain MG1363).
XX
XX Key Location/Qualifiers
XX CDS 178..1491
XX /tag= a
XX /product= Lysin.
XX complement (1542..1824)
XX /tag= b
XX
XX WO951561-A1.
XX
XX 23-NOV-1995.
XX
XX 12-MAY-1995; 95WO-NL00170.
XX
XX 12-MAY-1994; 94EP-0201353.
XX
XX (UNIL) QUEST INT BV.
XX
XX Buier G, Kok J, Ledebroer AM, Venema G,
XX WPI; 1996-010946/01.
XX P-PDB; AAR85285, AAR85288, AAR85289.
XX
XX Lysis of a culture of lactic acid bacteria in, e.g. cheese
XX production by in situ prodn. of an auto-lysin, regulated by an
XX inducible promoter.
XX
XX Claim 14; Page 51-54; 103pp; English.
XX
XX In situ production of a homologous autolysin or a heterologous
XX autolysin from a food grade Gram positive bacteria, can be used in
XX a process for the lysis of a culture of lactic acid bacteria. The
XX process can be used in the manufacture of products containing
XX cultures of lactic acid bacteria e.g. cheese, where the culture is
XX lysed following the completion of fermentation. The enhanced
XX induction of the autolysin is performed some hours after the
XX fermentation is finished. No extra lysin needs to be added and the
XX lysin does not need to be isolated or encapsulated. The time of
XX lysis can be precisely controlled.
XX
XX SQ Sequence 1929 BP; 619 A; 378 C; 321 G; 611 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.29 Length: 1929
XX Score: 130.50 Matches: 78
XX Percent Similarity: 37.93% Conservative: 43
XX Best Local Similarity: 24.45% Mismatches: 107
XX Query Match: 7.81% Indels: 92
XX DB: 17 Gaps: 15
XX
XX US-10-018-706-2 (1-322) x AA06135 (1-1929)
QY 26 ThrtHrCysIleuAlaGlyCysAlaSerLysProThrTYAsnSerThr-----42

Db 727 ACTGCTGCTTTACGGGAGATATGCTCCAGCCCAATGATGGCTTCAATGCC 786
QY 43 -----SergLysSerGlySerHisArgThrSergLys 52
Db 787 ATTATTTTCATATTAATTTGACTGTTTGACGAGGCTTCTTCCAGCTGGAATATCTAT 846
QY 53 SerGlyGlyLeuAlaIleGlySergLysValIleThrAspSergLysValProAsn--- 71
Db 847 TCTGGTGGC-----TCGACAAACCAATATTCGAATTAATTAATTCGAAACCAATGAC 897
QY 72 -----ArgTYrGlnValIleGlnIleAspThrValSeryIleAlaGlnArgTYr 88
Db 898 AGTTCAACTACTTAACCGTCGCAATCTGGTATCTCTTGGGGAATCTCACAAGATAT 957
QY 89 GlyLeuAsnTrpArgLysIleGlyHisIleAsnSnuLeuAsnSeryTYrThrIleTYr 108
Db 958 GGAATTTAGTGTGGCTCAAAATTCAAAGTCGATTAATCTTAAGATACC---ATTATCTAC 1014
QY 109 ThrGlyGlnTrpLeuThrLeuTrpSergLysAspLeuLysValArgGluArgSeryIle 126
Db 1015 ATTGCTCAAAAACCTGTACTGACAGGTTC-----GCTTCTTACAAATTC 1062
QY 129 SerGlyValAsnThrAlaHisThrProserProValAlaValGlnSerySergProPro 148
Db 1063 GGTGGTTCAAACAATTC-----GCAGAGCATCTACCAACC 1098
QY 149 ValGlnGlnHisProAlaValGlnLysProThrProProValAlaValIleLysArgPro 168
Db 1099 ACTTCTGTGACACTGCA-----AACCACTTCACAAACACTGTTAAGTTAATCC 1152
QY 169 ThrProThr-----ProProValValGlnGlnProAlaProValAlaProProValThr 186
Db 1153 GGAATATCCCTTTGGGCGCTTCAAGTAAATATTAATAACTAGTATGCTCAATGAAAGT 1212
QY 187 GluAlaProPheAlaThrGlySerySergLysValIleGln-----PheArgTYrPro 203
Db 1213 TGGAAATCATTTAAGTTCAGATGACATTTATGTCGCAAAATCTTATGTTTCACAATCT 1272
QY 204 ValGluAlaThrAsnProValAlaArgArgPheGlyThrAlaThrValAlaGlySeryThr 223
Db 1273 GCTGCTGCTTCAAAATCTCT-----TCGACAGGTTTCAGGCTCAACT 1311
QY 224 ValThrSeraSngLysMetTrpPheSergLysArgSergLysAspLeuIleAsnAlaSeraAsn 243
Db 1312 GCTACCAATTAAC-----TCAAAC 1329
QY 244 AlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSeryIleValIleGlnHis 263
Db 1330 TCG----- 1332
QY 264 ThrAsnGlyPheValSeryTYrIleHisIleLysAspAlaGlnValLysThrGlyAsp 283
Db 1333 ACTTCTTCTTACGCAAAATGCTCAATTCAT-----AGGTCGTTAAGGAT 1380
QY 284 ThrVal-----ArgThrGlyGlnArgGlyIleAsnMetLys-----Asn 296
Db 1381 ACTTCTGGGAGACTTGGCAAAATCTGCGACGCCAATTCCTTCAATCAAGGCTTGAAAT 1440
QY 297 GlnProSergLysAlaAlaLeuPheGluPheArgIleSeryArgAsnGlyValTYrVal 315
Db 1441 CATTTATCTTACGCACTACTAT-TTATTTGGTCAGTATCTCAAGATTAATAATTAATT 1496
RESULT 28
AAA23454
ID AAA23454 standard; cDNA; 4093 BP.
XX AAA23454;
XX
XX 19-JUN-2000 (first entry)
XX
XX cDNA encoding human secreted protein vb22_1, SEQ ID NO:63.
XX
XX
XX

CC from the vep gene cluster and module 7 from the Streptomyces ty1P.
CC gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC activity and can be used for PHA prodn. in host (esp. insect) cells
CC for use as a biodegradable polymer.

XX Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	16.4	15872	81	43	115	18
Percent Similarity:	130.50					
Best Local Similarity:	38.75%					
Query Match:	25.31%					
DB:	7.81%					

US-10-018-706-2 (1-322) x AAT68715 (1-15872)

```

Qy 34 AAlaserlypProthryrYasnSerThSerGlySerHisargThrSerGlySer 53
Db 3195 GCGTCGCGCCCTC-----TCGCGACGACGCCGCCACAGGTCTCTCGGGGAG 3151
Qy 54 -----GlyGlyLeuAlaIleGlySerGlnAlaIle 63
Db 3150 GTGACCGCCGCGGGGAGCGGACGATGCCATGATGCGATGCGCTCGCGGTG 3091
Qy 64 ThrAsp-----SerGlnGlyValProAsnArgTyrGlnValIleGlnIleAspThrVal 81
Db 3090 TCGCGCGGGGTCGCGGGGCGGATCCGTCGCGCCATCCGTCGCGCGCCGCCG 3011
Qy 82 SerIys-----IleAlaGlnArgTyrGlyLeuAsnArgTyrGlnIle 94
Db 3030 AGCAGGTCGCGCCAGTGGCGGCGGACGCGCGCGCTCGGCTGTCGAGACGATCCG 2971
Qy 95 ILGlyHisIleAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnIleThr 114
Db 2970 CTGGGCGCGCGACATCCGTCGTCGACGACGCGCTCGCGAC----- 2926
Qy 115 LeuTrpSerGlyAsp-----LeuIysValArgGluArgSer 126
Db 2925 -----TCGACGACATGAGGAGTCCAGCCGCTCTTGAAGTGAAGCCGATTCG 2872
Qy 127 IleSerSerGly-----ValAsnThrAlaHisThrProSerPro 139
Db 2871 ACGCGGTCGCGGTGTCGTCGTCGACGAGCGCGCGCGATGTCGCGACGCGTCG 2812
Qy 140 ValAlaValGlnSerSerArgProProValGlnIleHisProAlaValGlnIleProThr 159
Db 2811 GTCTCGTGGGAGCGCGCGCATCAACG-----CCACCCGACCGCGCGCTCG 2764
Qy 160 ProProValAlaValIleValIleProThrProThrProProValIleGlnIleProAla 179
Db 2763 CCC-----TCGCTCTCGCTCGCTCGCGCCGACGTCACACCCGCGCGCGG 2722
Qy 180 ProValAlaProProValIleGlnAlaProPheAlaThrGlySer----- 194
Db 2721 CCGGTGCGCGACCGGTGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2662
Qy 195 -----SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProValVal 211
Db 2661 CCGTCGACACGATGCGCTGCGCTGGAAGCGGATGTCGCGACGCGACCCCTGACCGT 2602
Qy 212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
Db 2601 CCGGTCTCCCGTCCGCGCGCGCTCAGTCAGTCGTCGCGCGCGCGCGCGCGCGCG 2542
Qy 226 Ser-----AsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db 2541 AGTCGCGCGACGACGATGCGGCTCGCGCGCGCGCG-----TTGCGCAGG 2497
Qy 244 AlaGlyThrValIleGlnAlaAsp-----HisAsnMetAspGlyAlaSerIleValIleGln 262
Db 2496 GCGGAGACGCGCGCGCTCTGTCGCGTACGAGATCCCGCGCGCATCGCGAGCAG 2437

```

```

Qy 263 HisThrAsnGlyPheValSerSerTyrIleHisIleIysAsp-----AlaGlnValIleThr 281
Db 2436 ACCCGTCGCGACCGACCTCCAGAGGTGTCGCGCGCGCGCTTCTCTCCAGGTGTCGTCG 2377
Qy 282 GlyAspThrValArgThrGlyGlnArgIleAlaSerMetIysAsnGlnProSerGlyAla 301
Db 2376 GCGTCGAGAGCGGTACCGGCGCTCGCG-----ACGTGTCACCCAGTACTCTCGGATCG 2323

RESULT 31
ID AA287283 standard; DNA, 15872 BP.
XX AA287283;
XX 05-JUN-2000 (first entry)
XX S. venezuelae vep ORF 1, SEQ ID NO:1.
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
XX neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
XX biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
XX chronic obstructive pulmonary disease; respiratory inflammation;
XX hypercholesterolemia; crop protection agent; ds.
XX Streptomyces venezuelae ATCC15439.
XX OS
XX FH Key Location/Qualifiers
XX CDS 20..13912
XX FT /*tag= a
XX FT /product= "vep ORF 1 amino acid sequence #1 (AAV77177)"
XX FT 14056..14151
XX FT /*tag= b
XX FT /product= "vep ORF 1 amino acid sequence #3 (AAV77199)"
XX FT 14167..15827
XX FT /*tag= c
XX FT /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"
XX FN WO200000620-A2.
XX PD 06-JAN-2000.
XX PE 25-JUN-1999; 99WO-US14398.
XX PR 26-JUN-1998; 98US-0105537.
XX PA (MINU ) UNIV MINNESOTA.
XX PI Sherman DH, Liu H, Xue Y, Zhao L;
XX DR WPI; 2000-160679/14.
XX RS P-PSDB; AAY77177, AAY77178, AAY77199.
XX PS Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX XX synthesis of methymycin and pikromycin -
XX XX Example 3; Figure 23; 438pp; English.
XX CC The invention relates to an isolated and purified nucleic acid segment
XX CC comprising a desosamine biosynthetic gene cluster, a fragment of its
XX CC biologically active variant, where the nucleic acid sequence is not
XX CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX CC streptomycetes antibiotics. The invention also relates to a macrolide
XX CC biosynthetic gene cluster, or fragments thereof. The macrolide
XX CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
XX CC pikromycin, neomethymycin, narbomycin or a combination of these
XX CC compounds. Recombinant or augmented cells comprising the desosamine
XX CC and/or macrolide biosynthetic gene clusters are useful for the
XX CC production of biologically active macrolides. The macrolide biosynthetic
XX CC proteins are useful for synthesis of methymycin, pikromycin,
XX CC neomethymycin and narbomycin. The alternative termination of polyketide
XX CC synthesis may be useful to prepare novel antibiotics and
XX CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
XX CC recombinant host cells are useful as biopolymers, e.g., in packaging or

```



```

US-10-018-706-2 (1-322) x ABA90521 (1-2365589)
Qy 3 ValThrIleAlaIleAsnSerGlnAsnGlnProIleValArgLeuGlyLeuIlePhe 22
Db 1065868 GTACCGG---ATTAACGCTACTGGCAATCTGGGCATAGGCATTTGGGCTGGT-TTC 1065923
Qy 23 GlyValIleThrThrCysIleLeuAlaGlyCysAla---SerLysProThrTyrAsnSer 41
Db 1065924 TCTTAACACAGCGCGCGCGCATTTACTCGGAATGACAGGGGATCTGATGCCAATCCA 1065983
Qy 42 ThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGln 61
Db 1065984 ACCGCTGATGAGGCGC-----GCGGACGCTCCTGATTTACGTTATGAGGGGTTTGG 1066031
Qy 62 ValIleThrAspSerGlnGlyValProAsnArg-----TyrGlnValLys----- 76
Db 1066032 CAATGACGAGATGATTCAGGGCTGACCTGACGCTGTTTATGATTAACCTCATGACA 1066091
Qy 77 -----GlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeu 90
Db 1066092 CGAGCAGAGTACGACGACCAATCCGACACATCA-----GCCCATTTCAACTCTTG 1066145
Qy 91 AsnTyrArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGly 110
Db 1066146 ATGTGG-----CATGCCAATAAC-----GGC 1066166
Qy 111 GlnThrLeuThrLeuThrSerGlyAspLeuLysValArgLysSerIleSerSerGly 130
Db 1066167 CAATGAGATTGGCAAAAGTTCTTATCTTATCTTGGACTCAATTCATGACATGACCAAT 1066226
Qy 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
Db 1066227 ATCATACTGCA-----ACGCAAGCTTTGTGACTACTTGAACGTCCTTTAAC 1066277
Qy 151 GlnHisProAla-----ValGlnLysProThrProProValValValLys 166
Db 1066278 GGAACACCTGAAACGTAGCACTTGGGCAACAGATGCTATCAAAATTTGTAACCTTAAA 1066337
Qy 167 LysProThrProThrProProValValGlnGlnProAlaProValAlaProProValThr 186
Db 1066338 ATCCCAACCGGTGGCGAGGTTATATTT-----GCTCCAAATTTCAAGTCTTATTCAC 1066388
Qy 187 GluLysProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAla 206
Db 1066389 -----GTRACAGAGCAATGGGT----- 1066406
Qy 207 ThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSer 226
Db 1066407 -----TGAGAACGAGTCCATCCAGCGGAGCGCAGAAATTTTAC 1066445
Qy 227 AsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsn----- 240
Db 1066446 AATGCTATG-----GACTTGTTATGTCATCCAACTCA 1066484
Qy 241 -----AlaSerAsnAlaGlyThrValIleGlnIleAlaAspHisAsn-----MetAsp 255
Db 1066485 ATCTTAGCTTTGCGAGATGCTCAAGTGTCCAGCGGAGAAATATATATGACTGGAT 1066544
Qy 256 GlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLys 275
Db 1066545 GGAATATTACACGGTCATCAACATGCGGATGATCTTATACAGGGTACGCACATCAAGC 1066604
Qy 276 AspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMet 294
Db 1066605 AGAATTCATGTTCTGTGGTCAAAATGTAAAGGCGCAACAAATGGAATTATG 1066661

```

```

RESULT 33
AI199683/c
ID AI199683 standard; DNA; 4403765 BP.
AC AI199683;
XX
XX
DT 15-JAN-2002 (first entry)

```

```

XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX KM variation; epidemiology; patient treatment; epidemic monitoring; db.
XX OS Mycobacterium tuberculosis.
XX US6294328-B1.
XX PM 25-SEP-2001.
XX PD 25-SEP-2001.
XX PF 24-JUN-1998; 98US-0103840.
XX PR 24-JUN-1998; 98US-0103840.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX DR WPI; 2001-647261/74.
XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX PT determining the nucleotide sequence of the strain at positions in the
XX PT genome corresponding to positions where M. tuberculosis strains CDC
XX PT 1551 and H37Rv differ.
XX PS Claim 4; SEQ ID NO 2; 3bp + Sequence Listing; English.
XX CC The invention relates to evaluating strain variation within and between
XX CC different populations of the tuberculosis bacterial pathogen.
XX CC Mycobacterium tuberculosis or related Mycobacterium by determining the
XX CC nucleotide sequence of the first strain at positions in the complete
XX CC sequence of the genome that correspond to positions that differ in the
XX CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AI199683) and
XX CC H37Rv (AI199682). The method is useful for evaluating strain variation of
XX CC M. tuberculosis and has valuable application in the fields of
XX CC tuberculosis genetics, epidemiology, patient treatment and epidemic
XX CC monitoring.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from USPTO
XX CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Alignment Scores:
Pred. No.: 1.46e+04 Length: 4403765
Score: 130.50 Matches: 81
Percent Similarity: 34.37% Conservative: 41
Best Local Similarity: 22.82% Mismatches: 128
Query Match: 7.81% Indels: 106
DB: 22 Gaps: 17

US-10-018-706-2 (1-322) x AI199683 (1-4403765)
Qy 40 AsnSerThrSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGly 59
Db 431307 AACTCGGCGACCGGCAACTGCGGTATCGGCACCTCGGACCGGCAACCGGCAATCGGC 431248
Qy 60 SerGlnValIleThrAsp-----SerGlnGlyValProAsnArgTyrGlnVal 75
Db 431247 AACACCGGACGACATTAACAGCGGCTTCTTCAACACCGGATCGTCACACCGGCTGCC 431188
Qy 76 LysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTyrArgGluIle 95
Db 431187 AACCGGCGACG-----TACCAACCGGCTGTATCAACAC 431152
Qy 96 GlyHisIleAsn----- 99
Db 431151 GGGCACACCAACACCGGATCGCAACCTGGGCGACTTCAACACGGGCTTCAACACCC 431092
Qy 100 ---AsnLeuAsnSerSerTyrThr-----IleTyrThrGlyGlnThrLeuThr 114

```

```

Db 431091 GGCATTTCAGTACGGGCTTTGGCCACAGAGGATATGCGCACCGGGGCTTTCATCACC 431032
QY 115 -----LeuTPsergLYAspLeuLYValAlaArgLIuArgSerIle 127
Db 431031 GGGCATGAGGGAACGGGCGCTTCTGGCGGGGAC-----CAGCAGGGCCATTTC 430981
QY 128 SerSerGIYValAsnThr-----AlaHisThrProSerProValAla 141
Db 430980 AGCGCGGCTATCGGGTCCATGATTCGCCGAATACCCGACAGCTCACCGTGAAGTTCCC 430921
QY 142 ValGln-SerSerArgProProValGlnGlnHisProAlaValGlnLYProThrProPr 161
Db 430920 GTCAACATCCCATCCATCCCGCGACT-----TCACCAACACCGATC 430882
QY 161 oValValVal-----ValLYserProThrProThrProProValValGlnGlnProAl 179
Db 430881 TACAGCGCATACCGCTTAGCAGCAATCACTTCGTTTACCATCCATCCAGATCCAGGATC 430822
QY 179 aProValAlaProProValThrGlnAlaProPhaAlaThrGlySerSerGIYValMetGI 199
Db 430821 CCCCTGCTGGCCGCGTCATCCAGCAAGCCGCTTCCCGCCATCACCGGACCGGTCGCC 430762
QY 199 nPheArgTYrProValGlyAlaThr-----AsnProValValArgArgPheGIYTh 216
Db 430761 GCGATCACCGTCACATCCGCGCGCCCTGGCGGTTTCAGCCGCG-ATCAGATCCCGGCGC 430703
QY 216 rAlaThrValAlaGlySerThrValThr----- 225
Db 430702 CCGAAGCGTCGCGCTTCGATGTCACGTTTCGTCACATTCGGCGGTACACAGGGGCTTTT 430643
QY 226 -----SerAsnGIYMetTrpPheSerGIYArgAspGIYAspLeu 239
Db 430642 CAACGCCACACCGATCCGTCCTCCGCGC---TTCTTCAACGCGCGCGCCGGAACCGTATC 430586
QY 239 eAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAspMetAspGIYAlaSerI 259
Db 430585 GGGCATCGGCACATCCGCGCGCCACATTCCTCGGCTTCGAGAGTGGCAACCTCGCG-- 430528
QY 259 eValIleGlnHisThrAsnGIYPhaValSerSerTYrIleHisIleLYAspAlaGlnVa 279
Db 430527 -----ACCTCGGGCTTC--AACACATACGCGCTCGCTCAATCGGGACTGCG 430484
QY 279 llysthrGIYAspThrVal-----ArgThrGIY-----GlnArgTI 291
Db 430483 GAACCTGGGCGATACCGTCTCGGCGGTATTCACAACCGGCATCGGGGACCGCCAAACGT 430424
QY 291 eAlaSerMetLYAsn----- 300
Db 430423 CTCGGGCATGTTCAACATCGGACGAGCAACCTCGCGGGGTTCTTCACAGCACGAGCGACGG 430364
QY 300 yAlaAlaLeuPheGlnPheArgIleSerArgAsnGIYValTYr 314
Db 430363 GATGTCGATGTTCAACCTCGCGCTGGGAAACATCGGCCAATTTC 430321

```

RESULT 34

AA199682/c

AA199682 standard; DNA, 4411529 BP.

AA199682;

15-JAN-2002 (first entry)

Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.

Mycobacterium tuberculosis, strain H37Rv, strain CDC 1551, genome;

variation; epidemiology; patient treatment; epidemic monitoring; de.

Mycobacterium tuberculosis.

US6294328-B1.

25-SEP-2001.

```

PF 24-JUN-1998; 98US-0103840.
PR 24-JUN-1998; 98US-0103840.
XX (GENO-) INST GENOMIC RES.
PA Fleischmann RD, White OR, Fraser CM, Venter JC;
PI WPI; 2001-647261/74.
XX
XX WPI; 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX determining the nucleotide sequence of the strain at positions in the
XX genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ
XX
XX Claim 3; SEQ ID NO 1; 3bp + Sequence Listing; English.
XX
XX The invention relates to evaluating strain variation within and between
XX different populations of the tuberculosis bacterial pathogen.
XX Mycobacterium tuberculosis or related Mycobacterium by determining the
XX nucleotide sequence of the first strain at positions in the complete
XX sequence of the genome that correspond to positions that differ in the
XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
XX H37Rv (AA199682). The method is useful for evaluating strain variation of
XX M. tuberculosis and has valuable application in the fields of
XX tuberculosis genetics, epidemiology, patient treatment and epidemic
XX monitoring.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from USPTO
XX at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 144602 G; 758379 T; 0 other;

Alignment Scores:
Pred. No.: 1.46e+04 Length: 4411529
Score: 130.50 Matches: 81
Percent Similarity: 34.37% Conservative: 41
Best Local Similarity: 22.82% Mismatches: 128
Query Match: 7.81% Indels: 106
DB: Gaps: 17

US-10-018-706-2 (1-322) x AA199682 (1-4411529)
QY 40 AsnSerThrSerGIYSerGIYSerHisArgThrSerGIYSerGIYLeuAlaIleGIY 59
Db 431224 AACTCGGGCGACCGGCACACTGGGGTATCGGCACATCCGCGACGCGCAACCGGCATCGGC 431165
QY 60 SerGlnValIleThrAsp-----SerGlnGIYValProAsnArgTYrGlnVal 75
Db 431164 AACACCGGCGACGACTAACACGCGGCTTCTTCAACACCGGCATCGTCACACCGCGTGGCC 431105
QY 76 lYsgInGIYAspThrValSerLYIleAlaGlnArgTYrGlyLeuAnThrArgGlnIle 95
Db 431104 AACCGGGCGAC-----TACAAACCGGCTGGTATACACACC 431069
QY 96 GlyHisIleAsp----- 99
Db 431068 GCGCAGACCAACACCGGATCGGCACACTGGCGGCACTTCACAGGGCTTACACACC 431009
QY 100 ---AsnLeuAnSerSerTYrThr-----IleTYrThrGIYGlnTrpLeuThr 114
Db 431008 GGCATTTTCATGACGGGCTTTCACCAACGAGGTATATCGCACCGGGGCTTTCATACCC 430949
QY 115 -----LeuTPserGIYAspLeuLYValArgGIYArgSerIle 127
Db 430948 GCGCAGATGGCAACGGCGCTTCTGGCGGGGAC-----CAGCAGGGCCATTATTC 430898
QY 128 SerSerGIYValAsnThr-----AlaHisThrProSerProValAla 141
Db 430897 AGCGCGGCTATCGGGTCCATGATTCGCCGAATACCCGACAGCTCACCGTGAAGTTCCC 430838
QY 142 ValGln-SerSerArgProProValGlnGlnHisProAlaValGlnLYProThrProPr 161

```

```

Db 430837 GTCAACATCCCATCCGACGCT-----TCACCAACACCGCTC 430799
Qy 161 ovalValVal1-----VallyLyseProThrProThrProValValGlnGlnProAl 179
Db 430798 TACAGCGGATACCGCTTACGACAAATCACTTCGTTTACACATGACATCGACGAGATC 430739
Qy 179 aProValAlaProValThrGlnAlaProPheAlaThrGlySerGlyValMetG1 199
Db 430738 CCCCTGCTGGCGGATGACATGACGACGCGCTTCCGCCCATCACCAGGACCGGTCC 430679
Qy 199 nPheArgTyrrProValGlyAlaThr-----AsnProValValAlaGlyPheGlyTh 216
Db 430678 GCGATCAGCGTCATCATCGGACCGCTTCGCGGTTCGACCGG-ATCAGATTCGCGGCAC 430620
Qy 216 rAlaThrValAlaGlySerThrValThr----- 225
Db 430619 CGCAAGCTCGGCTCCTTCATGTCACGTTGTCACATTCGCGGTACACGCGCTTTT 430560
Qy 226 -----SerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuI1 239
Db 430559 CAAGCCACACACGATCGCTCTCGGCG-TCCTTCAACGCGCGCCCGGACCGTATC 430503
Qy 239 eAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerI1 259
Db 430502 GGGCATCGCCCAACATCGCGCAATTCGCGCTTCAGAACGTCGCAACTCGCG- 430445
Qy 259 eValIleGlnHisThrAsnGlyPheValSerSerTrpIleHisIleGlyAspAlaGlnVa 279
Db 430444 -----ACCTCGGCGCTTC-----ACACATCAGCGCTCGCTGCAATCGGACCTGCG 430401
Qy 279 llyeThrGlyAspThrVal-----ArgThrGly-----GlnArgI1 291
Db 430400 GAACCTGGCGCATACCGCTCTCGGCGGTATTCACACCGGACATCGGCGCAACGCT 430341
Qy 291 eAlaSerMetLysAsn-----GlnProSerG1 300
Db 430340 CTCGCGCATGTTCAACATCGGACGACCACTCGCGGTTCTTCCACGACGACGCGACCG 430281
Qy 300 yAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyr 314
Db 430280 GATGTCATGTTCAACCTCGGCGCTGGCAACATCGGCCAATTC 430238

RESULT 35
AAL46625
ID AAL46625 standard; DNA; 1233 BP.
XX
AC AAL46625;
XX
DT 05-AUG-2002 (first entry)
XX
DE H influenzae BASB201 coding sequence #2.
XX
KW BASB201; otitis media; pneumonia; sinusitis; nosocomial infection;
KW auditive nerve damage; delayed speech learning; vaccine;
KW antibacterial; auditory; antiinflammatory; gene; ds.
XX
OS Haemophilus influenzae.
XX
FH Key Location/Qualifiers
FT CDS 1..1233
FT /tag= a
FT /product= "BASB201"
XX
XX MO200230967-A2.
XX
XX 18-APR-2002.
XX
XX 05-OCT-2001; 2001WO-EP11561.
XX
XX 13-OCT-2000; 2000GB-0025169.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX

```

```

PI 13; Page 87-88; 90pp; English.
XX
XX WPI: 2002-426267/45.
DR P-PSDB; MA017657.
XX
PT New isolated non-typeable Haemophilus influenzae BASB201 polypeptides,
PT useful as components of vaccines for treating bacterial infection such
PT as otitis media, delayed speech learning and inflammation of middle ear
PT
XX
XX Claim 13; Page 87-88; 90pp; English.
XX
XX The present invention provides the protein and coding sequences of
XX several versions of the BASB201 protein from non-typeable Haemophilus
XX influenzae. These can be used in the production of vaccines against H.
XX influenzae infection, which can cause otitis media in infants and
XX children, pneumonia in elders, sinusitis, nosocomial infections, or
XX invasive diseases, chronic otitis media with hearing loss, fluid
XX accumulation in the middle ear, auditive nerve damage, delayed speech
XX learning, infections of the upper respiratory tract and inflammation of
XX the middle ear. The present sequence is a version of the BASB201 coding
XX sequence of the invention.
XX
SQ Sequence 1233 BP; 463 A; 226 C; 241 G; 303 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0.818 Length: 1233
Score: 130.00 Matches: 64
Percent Similarity: 38.19% Conserved: 33
Best Local Similarity: 25.20% Mismatches: 111
Query Match: 7.78% Indels: 46
DB: 24 Gaps: 11
XX
US-10-018-706-2 (1-322) x AAL46625 (1-1233)
Qy 74 GlnValIleGlnGlyAspThrValSerIleAlaGlnArgTyrrGlyLeuAsnTrpArg 93
Db 562 CAATAAAACAAACAAACAGATTCACAAAGACAGACAGACATCAA----- 609
Qy 94 GlnIleGlnHisIleAsnLeuAsnSerSerTrpIleTyrrGlyGlnTrpLeu 113
Db 610 -----TCTAGCGTGAATGAATCAATTAATAATTAGCCCTGATCAAGATTA----- 657
Qy 114 ThrLeuTrpSerGlyAspLeuValArgGlnArgSerIleSerGlyValAsnThr 133
Db 658 -----TTGAATGCCCTTAAGCAACCAACAGACACTTCGTCAGAAATTCACAGA 708
Qy 134 AlaHisThrProSerProValAlaValAlaGlnSerSerArgProPro-----ValGln 150
Db 709 GCTGAA-----CAAGCAGTCGCGGACCAAGAAAGAGTGAAGAGAGGCACTTGCTCAA 762
Qy 151 GlnHisProAlaValGlnLysProThrProProValValValLysLysProThrPro 170
Db 763 CGCCAAAGAGTGAAGAAACGACAA----- 789
Qy 171 ThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPhe 190
Db 790 -----TCAAAACCTTATCAACCACTGTGCAAGAGCGCAATTA 828
Qy 191 AlaThrGlySerSerGlyVal-----MetGlnPheArgTyrrProValGlyAla 206
Db 829 CTTAATAGTACACGCGTTTACGGCGGCAAAAAACATATTCCTTACCAAGTTTCGGT 888
Qy 207 ThrAsnProValValAlaGlyPheGlyThrAlaThrValAlaGlySerThrValThSer 226
Db 889 TCA-----ATTTCGACTACTTTGGT-----TCTATCCAGACGCGAAGTCTGG 936
Qy 227 AsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr 246
Db 937 AAAGATGATGTAATTCGCGCATACGACGCGCTGTAAAGCAATTCCTGCTGAGCC 996
Qy 247 ValIleGlnAlaAspHis-----AsnMetAspGlyValaSerIleValIleGlnHisThrAsn 265

```


Qy 283 AsphThValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyValA 301
 Db 1177 GCCACGCTGAGGAAACACGCCATCGCCAAACCGGCCAACCCAGCGCGCG 1233
 RESULT 37
 AAX20250
 ID AAX20250 standard; DNA; 111309 BP.
 AC AAX20250;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE Borrelia burgdorferi polynucleotide sequence #3.
 XX
 KM Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KM infection; diagnosis; characterisation; detection; ds.
 XX
 OS Borrelia burgdorferi.
 PN MO9858943-A1.
 PD 30-DEC-1998.
 PF 18-JUN-1998; 98WO-US12764.
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MED-) MEDIMUNE INC.
 XX
 PI Clayton R, Dougherty BA, Fraser C, Iathigra R, Smith HO;
 PI White OR;
 DR WPI; 1999-081217/07.
 XX
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 XX
 PS Claim 1; Page 738-800; 1128pp; English.
 XX
 CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 XX
 SQ Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 10 other;
 Alignment Scores:
 Pred. No.: 243 Length: 111309
 Score: 128.50 Matches: 64
 Percent Similarity: 33.71% Conservative: 25
 Best Local Similarity: 24.24% Mismatches: 94
 Query Match: 7.69% Indels: 81
 DB: 20 Gaps: 9
 US-10-018-706-2 (1-322) x AAX20250 (1-111309)
 Qy 69 ValProAsnArg-----TyrGlnValLysGlnGlyAspThrValSerLysIle 84
 Db 45502 GTTCCCAATATGAAAGCAATGTTTATATTTGTTAAATAAATGACCTCATCTCTATA 45561
 Qy 85 AlaGlnArgTyrGlyLeuAsnTrpArgGluIleGlyHisIleAsnLeuAsnSerSer 104

Db 45562 GCTAGTCTTATATGATGTTCCCAAGGTGATATTTAGATTCTATATATCTTGAATGAA 45621
 Qy 105 TyrThrIleTyrThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeuValArgGlu 124
 Db 45622 --GTTTATTTTAAAGCAAAAGTGTATTTCTCGGGGAAAGTTG----- 45666
 Qy 125 ArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSer 144
 Db 45666 ----- 45666
 Qy 145 SerArgProProValGlnGlnHisProAlaValGlnLysProThrProProValVal 164
 Db 45667 -----CCCAAGAATTTTAAAGAG 45687
 Qy 165 ValLysLysProThr-----ProThrProProValValGlnGlnProAlaProVal 181
 Db 45688 GTATTAGGGGAGACTTTTATTTATTCCTGCGAGGCTTTATCTCGGGGTATGGCTAT 45747
 Qy 182 AlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArg 201
 Db 45748 CGACCAAGT-----CCGTTT-----ACAGAGTTATTAGTTTTCAC 45783
 Qy 202 TyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGly 221
 Db 45783 ----- 45783
 Qy 222 SerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAla 241
 Db 45784 -----AATGGAATGATATTCGCAATTTAGCTAATACGCCCTATTAAGCC 45828
 Qy 242 SerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyValAsn-----Ile 259
 Db 45829 TCAGAGAAAGGTGTTGTTGTTACTGACGATTTATTCGGAGGCTATGCAAAATATATTT 45888
 Qy 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnVal 279
 Db 45889 GTCATTTCTCAGACCAACGAGTTCCAAACTTATATGACACTTTGAATTCCTTTGGCGGT 45948
 Qy 280 LysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnPro--- 298
 Db 45949 AAGGTTGCAAAAGATTTCAGAGGCGGTATAGTTATAGGGAAGCACTGGCTAT 46008
 Qy 299 SerGlyValAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeu 318
 Db 46009 AGTACGGGCAATCATTTTCATTTCATTTTAAATGATAAAGTGAATAATCTATAG 46068
 Qy 319 ThrValLeuLys 322
 Db 46069 AATATTTTAAAGA 46080
 RESULT 38
 AAX20248
 ID AAX20248 standard; DNA; 910715 BP.
 AC AAX20248;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE Borrelia burgdorferi polynucleotide sequence #1.
 XX
 KM Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KM infection; diagnosis; characterisation; detection; ds.
 XX
 OS Borrelia burgdorferi.
 PN MO9858943-A1.
 PD 30-DEC-1998.
 PF 18-JUN-1998; 98WO-US12764.

PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MED1-) MEDIMUNE INC.
 XX
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
 PI White OR;
 DR WPI; 1999-081217/07.
 PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 PS Claim 1; Page 157-671; 1128pp; English.
 XX
 CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
 CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. *Borrelia* belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and *Borrelia* causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 XX
 SQ Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;
 Alignment Scores:
 Pred. No.: 3 07e+03 Length: 910715
 Score: 128.50 Matches: 64
 Percent Similarity: 33.71% Conservative: 25
 Best Local Similarity: 24.24% Mismatches: 94
 Query Match: 7.69% Indels: 81
 DB: 20 Gaps: 9
 US-10-018-706-2 (1-322) x AAX20248 (1-910715)
 QY 69 VALPROAEPARX-----TYRGLNVALLYEGGLYASPTHRVALSERLYSILE 84
 DB 634888 GTTCCCAATATGAAAGAAATGTTTATATGTTTAAAAAAGACCTCATCTCTATA 634947
 QY 85 AAGlnatgTYRGLYleuasntrpArgluilegLYHsileasnleuasnser 104
 DB 634948 GGTAGTGTCTATATGATGTTCCCAAGGTGATATTTAGATTCTATATATCTGATATGAA 635007
 QY 105 TYRThrileTYRThGLYlntrpLeuThrleuTrpserGLYAspleuLYsValArglu 124
 DB 635008 ---GTTTATTTTGGGCAAAAGTTGTTTATCCCGGGCAAGATTG----- 635052
 QY 125 ArgSerileSerSerGLYValenThrAlaHsThrProSerProVALAlaValGlnser 144
 DB 635052 ----- 635052
 QY 145 SerArgProProVALGlnGlnHisProAlaValGlnLYSProThProProVALValVal 164
 DB 635053 -----CCCAAGATTTTAAAAAGG 635073
 QY 165 VALlyslsYsProThr-----ProThrProProVALValGlnGlnProAlaProVAL 181
 DB 635074 GATTATGGGAGACCTTTTATTATTCCTGTCAGGGGTATTAATCTCGGGGTATGCTAT 635133
 QY 182 AlaProProVALThThGLNAlaProPhaAlaThrArglySerSerGLYValMetGlnPhArg 201
 DB 635134 CGACCAAGT-----CCGTTT-----ACAGGAGTATATGTTTGCAC 635169
 QY 202 TYRProVALGlyAlaThrAsnProVALArgArgPhaGlyThAlaThrValAlaGly 221
 DB 635169 ----- 635169

QY 222 SerThrValThSerAsnGLYMetTrpPheSerArgLYArgAspGLYAspleuileasnAla 241
 DB 635170 -----AATGGAAATATATGCAATTTAGCAATATACCCATTAAAGCC 635214
 QY 242 SerAsnAlaGLYThThValileGlnLAspHisAsnMetAspGLYAlaSer-----1le 259
 DB 635215 TCAAGAGAAGAGTGTGTGTGTATCTGACAGCATTTATAGCGGAGGATGAAATATATAT 635274
 QY 260 VALileGlnHisThrAsnGLYPhaValSerSerTYRleHisileYAspAlaGlnVal 279
 DB 635275 GTCAATTCTCACAGACGAGATCCAACTTATATGCAATTTGAATTCCTTGGCGGTT 635334
 QY 280 LYRThrGLYAspThrValArgThrGLYAlaArgileAlaSerMetLYAsnGlnPro--- 298
 DB 635335 AAGGTTGGAAAAAAGTTTCAAGGGACGGTATATAGTTATATGCAAGCATGCTAT 635394
 QY 299 SerGLYAlaAlaLeuPhaGluPhaArgileSerArgAsnGLYValTYRValAspProLeu 318
 DB 635395 AGTACGGGCAATCATTTGCAATTTTACATTTTAAAAATGTAAGAAATCTATATG 635454
 QY 319 ThrValleuLYs 322
 DB 635455 AATATTTAAGA 635466
 RESULT 39
 AAQ25083
 ID AAQ25083 standard; DNA; 1520 BP.
 XX
 AC AAQ25083;
 DT 15-NOV-1992 (first entry)
 XX
 DE Beta-lytic protease gene.
 XX
 KW Gram-negative; Gram-positive; bacteria; decomposition; ss.
 XX
 OS Achromobacter lyticus.
 XX
 FH Key Location/Qualifiers
 FT CDS 319..1443
 FT /tag=a
 XX
 JP04108387-A.
 XX
 PD 09-APR-1992.
 XX
 PF 29-AUG-1990; 90JP-0225136.
 XX
 PR 29-AUG-1990; 90JP-0225136.
 XX
 PA (WAKO) WAKO PURE CHEM IND.
 XX
 DR WPI; 1992-171653/21.
 DB P-PSDB; AAR24147.
 XX
 PT Beta-lytic protease gene and DNA encoding it - for decomposing
 PT Gram-positive and some Gram-negative bacteria
 XX
 PS Claim 1; Fig 1; 13pp; Japanese.
 CC Peptidase was purified from *Achromobacter lyticus* by a Sepharose
 CC CL-4B column, a Sephadex G-75 column, and by reverse phase HPLC.
 CC The first 25 N-terminal amino acids were found to be identical to
 CC those of the beta-lytic protease from *Lysobacter enzymogenes*. The
 CC beta-lytic protease gene of *A. lyticus* was cloned using PCR primers
 CC based on the sequence of the *l. enzymogenes* beta-protease. The
 CC full length sequence of the gene encoding the *A. lyticus* enzyme was
 CC determined by diideoxy sequencing. The beta-lytic protease is
 CC expected to be an enzyme which can decompose not only Gram-positive
 CC bacteria but also some Gram-negative bacteria.
 CC See also AAQ25084-6.
 XX
 SQ Sequence 1520 BP; 258 A; 533 C; 498 G; 231 T; 0 other;

Alignment Scores:

Pred. No.:	1.62	Length:	1520
Score:	127.50	Matches:	79
Percent Similarity:	33.70%	Conservative:	42
Best Local Similarity:	22.01%	Mismatches:	119
Query Match:	7.63%	Indels:	119
		Gaps:	15

US-10-018-706-2 (1-322) x AA025083 (1-1520)

```

Qy 10 GlnanglnlyspProilelyasrgleuGlyleuilephegilyvalilethrtThrcysalle 29
Db 313 CAGACCATGAGAGATTCCAGAGCGGACTG-----GGGCTGCGCTGCTGTCGCGC 366
Qy 30 leuilaaglycyasaserlyspProthrtYrzanseerthserglyserglysertharg 49
Db 367 CTGCGC-----ACGATCGCGCGGCGCAACGCGCGCGC 399
Qy 50 Thserglyserglylyleuallleeglyserglnvalilethraspsrglnlyval 69
Db 400 GCCACGCTCAG-----CGGCGAGGATCTGCTGATCTTCTACAGACGAGATG----- 444
Qy 70 ProanaaryrglnvallysglnGlyaspthrvalserlylealeglnargtyrly 89
Db 445 -----TTGCACTTCGACATCGATCGCATCGCCAGCAAGCATGCCGCGATCTGCAC 495
Qy 90 leuamntprarglulleglyhisileamamleuamanserlytrthletytrth 109
Db 496 AAGCACTCGGAAGATCTCGCAC----- 519
Qy 110 Glylntrpleuhtleuhtpserglyaspleuysvalarggluargserlyser 129
Db 520 -----TGCGCGCGC-----TACAGCGGATCTACCGCA 546
Qy 130 GlyvalantrhralahsthrproserProvalalavalglnserSerarproProval 149
Db 547 AGTGTGATCGCGCTGATGAGAGACAGAGCGCGGCTCAGCCAGCGCGCGACGATC 606
Qy 150 GlnlnhlsproalavalglnlyspProthrtProProvalvalvallylelyPro--- 168
Db 607 GTCGCTCGGACAGCTGCGCGCGCGCGCGC-----GCTTCGCGCGCGAGACCCCGC 657
Qy 169 -----ThrProthrtProProvalvalgln----- 176
Db 658 AGGTGCGCTGCGCGCTGCGCGAGTCGCTGTACAGACCGGATCCGACCGCCAGAGGCGC 717
Qy 176 ----- 176
Db 718 GGTGACGTGCGCGCGCCATCCGCTGTCAGAGCGTTCCGCGGACCAACGA 777
Qy 177 -----GlnProalaprovalalaproProvalthr 186
Db 778 GCCGCGCGCGCTGCGCGCGAGTCGAGTTCAGTCTGCTACAGCGCCCGCTGTTCA 837
Qy 187 Glu---Alapro-----Phe 190
Db 838 CGAACCGCGCGCGCAAGCGCGCTTCGACCGCTTCGCCAAGCGCGCGCGACGTGCA 897
Qy 191 AlaThrglyserSerglyvalmetGlnpheargtyrProvalalathraspProval 210
Db 898 GCCGCTGTCGCGCGCGCTGCTGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 957
Qy 211 ValargargpneglythralaThrvalaAaglyser----- 222
Db 958 -----GGCGCGCGCGCGCAACCAACCGCTTCGCAATTCACCGCATGTCGCTG 1008
Qy 223 ThrValThrseranglymetrPheSerglyargaspGlyaspleuileasnalaser 242
Db 1009 GACATGTCGCGCGCGCGCTGCGCGCGCAACCAACCGCATGTCGCTGCTGCTG 1068
Qy 243 AsnlaaglythralilleglnlaasphlsanmetaspGlyalaserilevalillegln 262

```

```

Db 1069 GCCCGCGCTGCTCAAGCGC-----CACTCTTCGTCCTTCGCGAGATGCTG----- 1116
Qy 263 HsthranglyphevalaserSerTyrllehisilelyaspaalaglnvallythrgly 282
Db 1117 CACACCGCGCGCTGCTGACGACCTTACCACTGATGACATCAGTACACACCGCGC 1176
Qy 283 AspThrvalargthrglnarglaleasermetlysaanglnproserglyala 301
Db 1177 GCCACGTGTGATGAGAACCGGCATCGCCACCGCGCGCAACACCGCGCGCGCGC 1233

RESULT 40
AAZ00266
ID AAZ00266 standard; DNA; 7277 BP.
XX
AC AAZ00266;
XX
DT 12-OCT-1999 (first entry)
XX
DE HEV-US2 full length gene sequence.
XX
KW Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
XX vaccine; passive immunisation; ss.
XX
OS Hepatitis E virus.
XX
PN MO9919732-A1.
XX
PD 22-APR-1999.
XX
PF 15-OCT-1998; 98WO-US21941.
XX
PR 15-OCT-1997; 97US-0061199.
XX
PA (ABBO) ABBOTT LAB.
PI Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
DR WPI; 1999-288017/24.
XX
PT Detection of United States isolates of hepatitis E virus
XX
PS Claim 36; Page 201-204; 260pp; English.
XX
CC The invention provides a method for detecting a US (sub)type hepatitis E
CC virus (US-HEV), or its naturally occurring variants in a sample by
CC treatment with a binding partner specific for a marker of the virus, and
CC then detecting any complex formed. The method is used to diagnose
CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC are useful in vaccines or for passive immunisation. The polypeptides are
CC also used to raise specific antibodies (useful as immunoassay reagents).
CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC usual hybridisation and amplification assays for detecting infection. The
CC present sequence represents a HEV-US2 full length gene sequence.
XX
SQ Sequence 7277 BP; 1333 A; 2111 C; 1904 G; 1916 T; 13 other;

Alignment Scores:
Pred. No. 11.7 Length: 7277
Score: 127.00 Matches: 80
Percent Similarity: 33.14% Conservative: 37
Best Local Similarity: 22.66% Mismatches: 120
Query Match: 7.60% Indels: 116
Gaps: 19

US-10-018-706-2 (1-322) x AAZ00266 (1-7277)

Qy 27 Thrvalleuamlaaglycyas-----AlaserlyspProthrtYrzanseerThrs 43
Db 1755 ACGACGTGTGATGAGCGCCCATCTTGAAGCGATGCCCTGAGAGATGTTCTGTCA 1814
Qy 44 Glyserlyserhlsargthrsrglyserglylyleuallale----- 58

```

```
Db 1815 TTGACGCTCTGCGAGCTATATGAGGCGGGTCGACAGCCTCACTTATGAGCTCACC 1874
QY 59 -----GlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrglnValLys 76
Db 1875 CCGGCGGCTGTCGAGGTAAAGATTTCATCTAATGCTGTGATTCACATGCCACATTCGCC 1934
QY 77 GlnGlyAspThrValSer----- 82
Db 1935 CCGGTGGGCGCCCTAGCCCGCGGGGAGGTGGCGSCTTCGCAGTGCTCTTAT 1994
QY 83 -----LysIleAlaGlnArgTyrglyLeuAsn-----TrpArg----- 93
Db 1995 AGATACATATAGTTCACCCAGCGGCACTTGCCTGACGCGGACTATGGCTACATCCTGAG 2054
QY 94 -----GluIleGlyHisIle-----AsnAsn 100
Db 2055 GGGCTGCTGGGTATCTTCCCCCATTTCTCCCTGGGCATATTTGGAGTCTGCTAACCCC 2114
QY 101 LeuAsnSerSerTyrglyThrIleTyrglyGlnTrpLeuThrLeu---TrpSerGlyAsp 119
Db 2115 TTTTGGGAGAGGGGCTTTGTATACCCGACCTGTCAACCTGTGTTTCTAGTGAAT 2174
QY 120 LeuLysValArgGlnArgSerIleSerSerGlyValAsnThrAlaHisThrProSerPro 139
Db 2175 TTCTCCCCCTGAGCGCGGCTCTCTCGCTGCGCTGCCGCCCGGGGTTGCCCTACCT 2234
QY 140 ValAlaValAlaGlnSerSerArgProProVal----- 149
Db 2235 ACT-----CCACCTGTAGATATCTGGGTTTACACGCGCCTCA 2276
QY 150 ---GlnGlnHis-----ProAlaValGlnLysPro-----Thr 159
Db 2277 GAGGAATCTCATGTGTATGCGGAGCTGATCTGCTGTTCTGAGCTGCTGATTTGACC 2336
QY 160 ProProValValValValLysProThrProThrProProValValGlnGlnProAla 179
Db 2337 AGCCCTATGTGCTTAC-----CCCCCCCCCTCTCCCTGCGTAAGCGGCA 2390
QY 180 ProValAlaProProValThrGlnLysProPheAlaThrGlySerSerGlyValMetGln 199
Db 2391 ACATCCCGCTCCCGGCACTGCGCT-----CTCCTT 2423
QY 200 PheArgTyrglyProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrVal 219
Db 2424 TACACCTACCCCGACGCGGCC-----AAGGTGTAAT 2453
QY 220 AlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIle 239
Db 2454 GCGGGGTCAATTGTGTGAGTCAGACTGTGATTGG-----TTAGTC 2492
QY 240 AsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259
Db 2493 AATGCTCAAAACCCCTGGC-----CATGCCCGCGGGGTGGC----- 2528
QY 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrglyIleHisIleLysAspAlaGlnVal 279
Db 2529 ---CTTGGCAT-----GCTTTTATCAAGTTTCCAGAAAGCGTTTAC 2570
QY 280 LysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnPro--- 298
Db 2571 TCGACTGAATTCATCATGCGGAGGCGCTTCGACGATACACTTAAACCCGCGCCTAAT 2630
QY 299 SerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsn 311
Db 2631 ATCCATGCAAGTGGCTCCGACTATAGGTTGAGCAAAAC 2669
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 3, 2003, 20:12:12 ; Search time 48 Seconds
(without alignments)
2057.290 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 1670
Sequence: 1 MVTAINSONQKPIKRLGH.....LFEFRISRGVYDPLTVLK 322

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=x1h
-Q=/cgn2_1/USFTO.spool/US10018706/runat_30062003_091105_23816/app_query.fasta_1.519
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=smi -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=ppc -THR MAX=100 -THR MTN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pclo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10018706 @CNC 1.1 32 @runat_30062003_091105_23816 -NCPU=6 -ICPU=3
-NO MAP -LARGEOBURY -NEG_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG
-DELT=TIMEOUT=120 -WASN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219.5	13.1	1215	5	PCT-US96-05320A-707
2	209	12.5	2179	4	US-09-405-728-4
3	165	9.9	1885	3	US-08-619-812-3
4	131	7.8	3646	3	US-08-619-812-7
5	130.5	7.8	1930	2	US-08-737-716-1
6	130.5	7.8	1930	2	US-08-737-716-11
7	130.5	7.8	15872	4	US-09-105-537-1
8	130.5	7.8	4403765	4	US-09-103-840A-2
9	130.5	7.8	4411529	4	US-09-103-840A-1
10	123.5	7.4	4411529	4	US-09-103-840A-1
11	120.5	7.2	3004	4	US-08-276-213-6
12	119.5	7.2	11707	4	US-09-136-574A-1

13	118.5	7.1	4403765	4	US-09-103-840A-2	Sequence 2, Appli
14	118	7.1	5127	4	US-09-462-606-1	Sequence 1, Appli
15	118	7.1	7207	4	US-09-462-606-3	Sequence 3, Appli
16	117.5	7.0	2830	2	US-09-010-928B-1	Sequence 1, Appli
17	116	6.9	4518	3	US-09-125-287-2	Sequence 2, Appli
18	116	6.9	12839	3	US-09-125-287-1	Sequence 1, Appli
19	115	6.9	3300	1	US-08-194-290-6	Sequence 6, Appli
20	115	6.9	3300	2	US-08-614-377A-6	Sequence 6, Appli
21	115	6.9	3300	4	US-09-142-648B-6	Sequence 6, Appli
22	113.5	6.8	1896	4	US-09-311-626B-15	Sequence 15, Appli
23	113.5	6.8	7100	4	US-09-308-375-1	Sequence 1, Appli
24	112.5	6.7	2186	2	US-08-878-546-9	Sequence 9, Appli
25	112.5	6.7	8438	1	US-07-945-283-1	Sequence 1, Appli
26	112	6.7	36519	3	US-08-923-137-2	Sequence 2, Appli
27	111	6.6	867	4	US-09-071-035-221	Sequence 221, App
28	111	6.6	993	4	US-09-134-001C-1165	Sequence 1165, Ap
29	110.5	6.6	2077	1	US-08-217-327-7	Sequence 7, Appli
30	110.5	6.6	6416	6	US-09-136-574A-2	Sequence 2, Appli
31	109.5	6.6	2029	4	US-09-136-574A-46	Sequence 46, Appli
32	109	6.5	3833	1	US-08-917-320-18	Sequence 18, Appli
33	109	6.5	3833	5	PCT-US95-04611A-18	Sequence 18, Appli
34	109	6.5	5931	3	US-08-783-774-1	Sequence 1, Appli
35	109	6.5	5931	4	US-09-556-706B-1	Sequence 1, Appli
36	108	6.5	3172	1	US-08-314-309A-1	Sequence 1, Appli
37	107.5	6.4	2061	2	US-09-258-371-9	Sequence 9, Appli
38	107.5	6.4	2061	3	US-08-751-230-9	Sequence 9, Appli
39	107.5	6.4	2061	3	US-09-499-082-9	Sequence 9, Appli
40	107.5	6.4	2061	4	US-09-258-372-9	Sequence 9, Appli
41	107.5	6.4	2061	4	US-09-006-783A-2	Sequence 2, Appli
42	107	6.4	1470	4	US-09-109-841-1	Sequence 1, Appli
43	106.5	6.4	1446	4	US-09-149-476-139	Sequence 139, App
44	106.5	6.4	8367	2	US-08-583-562B-7	Sequence 7, Appli
45	106.5	6.4	8367	2	US-08-779-113-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
PCT-US96-05320A-707
Sequence 707, Application PC/TUS9605320A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences
APPLICANT: 9410 Key West Avenue
APPLICANT: Rockville, MD 20850
APPLICANT: United States of America
APPLICANT: Johns Hopkins University
APPLICANT: 720 Rutland Avenue
APPLICANT: Baltimore, MD 21205
APPLICANT: United States of America
APPLICANT: Mark D. Adams
APPLICANT: Owen White
APPLICANT: Hamilton O. Smith
APPLICANT: J. Craig Venter
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20003-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April 22, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102

FILING DATE: June 7, 1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/487,429
 FILING DATE: June 7, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Eric K. Steffe
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.014PC01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2540
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 707:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1215 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 PCT-US96-05320A-707

Alignment Scores:
 Pred. No.: 5,08e-12 Length: 1215
 Score: 219.50 Matches: 76
 Percent Similarity: 44.63% Conservative: 46
 Best Local Similarity: 27.84% Mismatches: 114
 Query Match: 13.14% Indels: 37
 DB: Gaps: 10

US-10-018-706-2 (1-322) x PCT-US96-05320A-707 (1-1215)
 QY 71 AANAATGTYGlnValIysGlnIyAspThrValIserIySleIaGlnArGTYrGlyLeu 90
 Db 433 AATACCTTAATAAGATAAACAAGCGATACATGTTCTTAATGCTTACTGCGCGGAGATT 492
 QY 91 AANTPARGluIleGlyHisIleAsnLeuAsnSerIyThrIleTyThrGly 110
 Db 493 GATGTAAAGAATTCGCGCGCTTGATTAACCTATCCGAACCTTAATTAAGTTAGCG 552
 QY 111 GlnTPLeuThrIleuTrpSerGlyAspLeuIyValArgGluArgSerIleSerGly 130
 Db 553 CAAGTTTAAAGATTCTTAATGATGACATTAACGTTACTACACAGTTCT 606
 QY 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
 Db 607 -----GTAAACAACCTCGACATGACATCAACGACATCAACGACATTA 651
 QY 151 GlnHisProAlaValGlnIySerProThrPro----- 160
 Db 652 -----CAGCACTACTTAT---ACTCCGGGGCGCAACCGTACTCAATATGATCTGAC 702
 QY 161 -----ProValValValIySlyProThrProThrProThrProProValVal 175
 Db 703 GGTACTATTATTTGGCCCAATTAATCAGAGCGTGGACATCGCCCTAGTACTCTGTGCA 762
 QY 176 GlnGlnProAlaProValAlaProProValThrGluAlaProPheAlaThrGlySerSer 195
 Db 763 ACTTCATCAACAACAGGTATCATCTGTAAATAATGCAATAGCACAACCAATTAAATTC 822
 QY 196 GlyValMetGlnPheArgTYrProValGlyAlaThrAsnProValAlaArgArgPheGly 215
 Db 823 AATGCTGTTGGG-----CCAAATC---GCCCAAAAT---GTTGTGTGGCAATGGCG 867
 QY 216 ThrAla-----ThrValAlaGlySerThrValThr-----SerAsnGlyMetTrp 230
 Db 868 ACTTCAGGTATATATCATCAAGGTTTCTCAAGCAGACAGATGGCGGTAACAAAGGAATGAT 927
 QY 231 PheSerGlyArgArgGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla 250
 Db 928 ATTAGTGGTTCACAGGACAGAGCTGTAAAGCGGCTGACAGGCGCAATGCTGTATGCA 987
 QY 251 AspHisAsnMetArgGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheVal 268
 Db 988 GGCAATGCTTACGTTAGCTTACGTAATTTATATTCATCAACATATATGATGATTTTITA 1047

QY 269 SerSerTYIleHisIleLeuAspAlaGlnValIySerThrGlyAspThrValArgThrGly 288
 Db 1048 AGTGCCTATGGCGATACAGCAAAATCTTCTCCCGCATCAACAAGATCCAAAGCAGGT 1107
 QY 289 GlnArgIleAlaSerMetIyAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIle 308
 Db 1108 CAAGCATTCGCAAAAATGGTGGTGGCTGCTGTAACAATACCGTGAATCTCCACTTGAAATT 1167
 QY 309 SerArgAsnGlyValTYrValAspProLeuThrValIleu 321
 Db 1168 CGCTATAAGGTAAATCATGTTGATCGATCGATTACTTAACCTA 1206

RESULT 2
 US-09-405-728-4
 Sequence 4, Application US/09405728
 Patent No. 6391316
 GENERAL INFORMATION:
 APPLICANT: Potler, Andrew A.
 APPLICANT: Rioux, Clement
 APPLICANT: Schryvers, Anthony B.
 TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
 TITLE OF INVENTION: TRANSFERIN-BINDING PROTEINS
 FILE REFERENCE: 9000-0049.20
 CURRENT FILING DATE: US/09/405,728
 EARLIER FILING DATE: 1999-09-24
 EARLIER APPLICATION NUMBER: US 09/267,749
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 4
 LENGTH: 2179
 TYPE: DNA
 ORGANISM: Haemophilus somnus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (872)..(1906)
 US-09-405-728-4

Alignment Scores:
 Pred. No.: 1.3e-10 Length: 2179
 Score: 209.00 Matches: 63
 Percent Similarity: 44.09% Conservative: 49
 Best Local Similarity: 24.80% Mismatches: 106
 Query Match: 12.51% Indels: 36
 DB: Gaps: 6

US-10-018-706-2 (1-322) x US-09-405-728-4 (1-2179)
 QY 73 TYGlnValIySglnIyAspThrValIserIySleIaGlnArGTYrGlyLeuAsnTrp 92
 Db 1229 TACAAGTACGCAAAAGCGATACATGTTCTTAATGCTTAATTCAGGCATGATATA 1288
 QY 93 ArgGluIleGlyHisIleAsnLeuAsnSerSerTYrThrIleTyThrGlyGlnTrp 112
 Db 1289 AAGAAATGGCCACACATTAATATATGCTGACACCATATCATCTGATATGGACAAGTA 1348
 QY 113 LeuThrIleuTrpSerGly-----AspLeuIyValArgGluArgSerIleSerSer 129
 Db 1349 TTGAATATGCAAAATATATTCGCCGATGACATATGATACCAACACACATTAATGAA 1408
 QY 130 GlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProVal 149
 Db 1409 TCGAGGTGACCAAAATATACGTCATGAG--ACATGAAATGCTATTAATTAACCAAT 1465
 QY 150 GlnGlnHisProAlaValGlnIySProThrProProValValValIySlySlyProThr 169
 Db 1466 GACCAATATGAACCCGTTGCTACACACACATTCACCAATGCGCAATCAATTAATA 1519
 QY 170 ProThrProProValValGlnGlnProAlaProValAlaProProValThrGluAlaPro 189
 Db 1520 ---ACACCTCCAGCCACCTCAATATATATGATGATTTGGCA----- 1558
 QY 190 PheAlaThrGlySerSerGlyValMetGlnPheArgTYrProValGlyAlaThrAsnPro 209

Db 1559 -----ACAAATGGAAAA 1570
Qy 210 ValValaArgArpGlyThrAlaThrValAlaGlySerThrValThSerAsnGlyMet 229
Db 1571 ATTATTCAGAGATTTCAGGCTGATGAGGC-----AATTAAGGTATT 1615
Qy 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGln 249
Db 1616 GATATTACGGTTCCTCGGACCAAGCTGTTAATGACAGCGCTGGACGAGTTGTATAT 1675
Qy 250 AlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPhe 267
Db 1676 GCCCGAGAGCGCTTACGTGATGATGATATTATTAATTAATTAACATATGACAGTTAT 1735
Qy 268 ValSerSerGlyThrIleHisIleLeuAspAlaGlnValIleGlyThrGlyAspThrValArgThr 287
Db 1736 TTAAGTGTATTAGCATTAATGAAGTAACTGCTCAAGATCAGCAAGAAAGTTAAACG 1795
Qy 288 GlyGlnArgIleAlaSerMetLeuAsnGlnProSerGlyAlaAlaLeuPheGlnPheArg 307
Db 1796 GGTCAACAATTTGCTTAATGGAATGGAAGTTCTGGAACAACATCAACTCCATTGTGA 1855
Qy 308 IleSerArgAsnGlyValIleValAspProLeuThrValIleu 321
Db 1856 ATTCGTTTAAAGTCAATCAGTAGATCCATGAGATATTTA 1897
RESULT 3
US-08-619-812-3
; Sequence 3, Application US/08619812
; Patent No. 6100066
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: THEISEN, MICHAEL
; APPLICANT: HARLAND, RICHARD J.
; APPLICANT: KIDOU, CLEMENT R.
; TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBIN
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619, 812
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038, 719
; FILING DATE: 29-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0019, 20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS

LOCATION: 872..1708
US-08-619-812-3
Alignment Scores:
Pred. No.: 1.85e-06 Length: 1885
Score: 165.00 Matches: 57
Percent Similarity: 43.80% Conservative: 49
Best Local Similarity: 23.55% Mismatches: 98
Query Match: 9.88% Indels: 38
DB: 3 Gaps: 6
US-10-018-706-2 (1-322) x US-08-619-812-3 (1-1885)
Qy 73 TyrGlnValIleGlySerGlyAspThrValSerIleAlaGlnArgGlyLeuAsnTrp 92
Db 1229 TACAAGTACCAAGCAAGGAGATACATGTTCTTATGCTTATATTCAGCATGATATA 1288
Qy 93 ArgGluIleGlyHisIleAsnAsnLeuAsnSerGlyThrIleTyrThrGlyGlnTrp 112
Db 1289 AAAGATTGGCCACACTTAATATATATGCTGAGCCATATCATCTGATGATGACACAGTA 1348
Qy 113 LeuThrIleuTrpSerGly-----AspLeuYsValArgGluArgSerIleSerSer 129
Db 1349 TTGAATAATTGCAATTAATATATCCCATAGCAATATGATGATGACACACACATTAATGA 1408
Qy 130 GlyValAsnThrAlaHisThrProSerProValAlaValAlaGlnSerSerArgProProVal 149
Db 1409 TCAGAGGTGACACAAATATACGTCAATGAG--ACATGATATGCTTAATTAACCAACAAT 1465
Qy 150 GlnGlnHisProAlaValAlaGlnLysProThrProProValAlaValAlaLysProThr 169
Db 1466 GAACAATAAGAAACCGTTGCTACACCAACATTCACACATGACCAATCAATATAA----- 1519
Qy 170 ProThrProProValAlaGlnGlnProAlaProValAlaProProValThrGlnAlaPro 189
Db 1520 --ACACTTCAGCCACCTCAATATATAGCTTGATTGGCCA----- 1558
Qy 190 PheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValAlaGlyAlaThrAsnPro 209
Db 1559 -----ACAAATGGAAAA 1570
Qy 210 ValValaArgArpGlyThrAlaThrValAlaGlySerThrValThSerAsnGlyMet 229
Db 1571 ATTATTCAGAGATTTCAGGCTGATGAGGC-----AATTAAGGTATT 1615
Qy 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr--ValIleG 249
Db 1616 GATATTACGGTTCCTCGGACCAAGCTGTTAATGACAGCGCTGGACGAGTTGTAT 1675
Qy 249 IlaIaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyP 267
Db 1676 ATCCCGAGAGCGCTTACGTGATGATGATATTATTAATTAATTAACATATGACAGTT 1735
Qy 267 heValSerSerGlyThrIleHisIleLeuAspAlaGlnValIleGlyThrGlyAspThrValArgT 287
Db 1736 ATTTAAGTGTATTAGCATTAATGAAGTAACTGCTCAAGATCAGCAAGAAAGTTAAAG 1795
Qy 287 hrGlyGlnArgIleAlaSerMetLeuAsnGlnProSerGlyAlaAlaLeuPheGlnPhe 307
Db 1796 CGGGTCAACAATTTGCTTAATGGAAGTTCTGGAACAACATCAACTCCATTGTTA 1855
Qy 307 rg 307
Db 1856 AA 1857
RESULT 4
US-08-619-812-7
; Sequence 7, Application US/08619812
; Patent No. 6100066
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: THEISEN, MICHAEL
; APPLICANT: HARLAND, RICHARD J.

APPLICANT: RIOUX, CLEMENT R.
TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: REED & ROBINS
STREET: 635 BRYANT STREET
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,812
FILING DATE: 15-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/038,719
FILING DATE: 29-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9000-0019.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3231
TELEFAX: (415) 617-8999
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3646 base pairs
TYPE: nucleic acid
TOPOLOGY: linear
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2772
FEATURE:
NAME/KEY: CDS
LOCATION: 2776..3570
US-08-619-812-7
Alignment Scores:
Pred. No.: 0.00986 Length: 3646
Score: 131.00 Matches: 48
Percent Similarity: 43.75% Conservative: 43
Best Local Similarity: 23.08% Mismatches: 79
Query Match: 7.84% Indels: 38
DB: 3 Gaps: 6
US-10-018-706-2 (1-322) x US-08-619-812-7 (1-3646)
QY 73 TTYGlnValIvGlnGlyAspThrValSerIySILaIaGlnAryrGlyLeuAentp 92
DB 3091 TCAAGATGCGAAGGAGGATCCATGCTTCTTATGCTTATTTTCAGCGATGATATA 3150
QY 93 ATGCGuileGlyHisIleAsnLeuAenSerSerTyThrIleYrThrGlyGlnTTP 112
DB 3151 AAGAGATTGGCCACTAATATATAGTGTGAGCCATTCATTCAGTATTTGACAGAGTA 3210
QY 113 LeuThrIeUTrpSerGly-----AspLeuIvValIaGlnAryrSerIleSerSer 129
DB 3211 TTGAATAATTGCAATATATATTTCCCGATGCAATATGATACCAACAGCAATTAATGTA 3270
QY 130 GlyValAenThrAlaHisThrProSerProValAlaValGlnSerSerArgProVal 149
DB 3271 TCAGAGGTGACCAAAATACAGTCAATAG--ACATGGAATCTTATTAACCAACAAT 3327
QY 150 GlnGlnHisProlAlaValGlnIvSProThrProProValValValValIySProThr 169
DB 3328 GAACAAATGAAACCCGTGTCTACACCAACACATTCAAACATGCAATCAATATA----- 3381

QY 170 ProThrProProValValGlnIvSProThrProProValAlaProProValThrGlnAlaPro 189
DB 3382 ---ACACCTCCAGCCACCTCAATATATGCTTGATTTGGCCA----- 3420
QY 190 PheAlaThrGlySerSerGlyValIvGlnPheAlaGlyrProValGlyAlaThrAsnPro 209
DB 3421 -----ACAAATGCAAAA 3432
QY 210 ValValIaGlyArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
DB 3433 ATTATTCAAGATTTTCCAGTCTGATGAGCC-----AATTAAGTATT 3477
QY 230 TTPPheSerGlyIaGlyAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr--ValIleG 249
DB 3478 GATATTAGCGGTTCTCGTGAGACAGCTGTTATATGACAGCTGATGAGCGAGTGTAT 3537
QY 249 IValIaAspHisAsnIleAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyP 267
DB 3538 ATGCCGAGACCGCTTACGATATGATATTATTAATTATTAACATATATGACAGT 3597
QY 267 heValSerSerTyThrIleHis 273
DB 3598 ATTTAAGTCTTATGACAT 3617
RESULT 5
US-08-737-716-1
Sequence 1, Application US/08737716
Patent No. 5955258
GENERAL INFORMATION:
APPLICANT: GIBBE, BURST
APPLICANT: GERARD VENEMA
APPLICANT: USN KOK
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/NL95/00170
FILING DATE: 12-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94201353.3
FILING DATE: 12-MAY-1994
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
STRAIN: MG1363
IMMEDIATE SOURCE:
CLONE: Fig.4 (publ.1)-acma and ORFA
FEATURE:

```

NAME/KEY: CDS
LOCATION: 178..1489
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1542..1824)
US-08-737-716-1

Alignment Scores:
Pred. No.: 0.00412 Length: 1930
Score: 130.50 Matches: 73
Percent Similarity: 37.71% Conservative: 39
Best Local Similarity: 24.58% Mismatches: 96
Query Match: 7.81% Indels: 89
DB: Gaps: 14

US-10-018-706-2 (1-322) x US-08-737-716-1 (1-1930)
QY 26 ThrThrcysileuallaglycysalaserlyserprothryraenserthr----- 42
Db 727 ACTGCTGCTTTGACGGGAGATATGCTACCGACCAAGTATGCGCTTCACTGAATCGC 786
QY 43 -----SerGlySerGlySerHisArgThrSerGly 52
Db 787 ATTAATTTCTCAATATAATTGACTCGTTTGACGAGCTTCTTCACTGGAATATCTAAT 846
QY 53 SerGlyGlyLeuAlaileGlySerGlnValileThrAspSerGlnGlyValProAsn--- 71
Db 847 TCTGTGTGC-----TCGACACACCAATTTACGATATATATTCGACCAACATATGC 897
QY 72 -----ArgTyrglnVallyserGlnGlyAspThrValSerlySileAlaGlnArgTy 88
Db 898 AGTTCACACTACTATACCGCTCAATCTGTGATCTCTTGGGAATCTCAAAAGATAT 957
QY 89 GlyLeuAsnTPraGlnGlnleGlyHisileAsnleuAsnSerSerTyThrIleTy 108
Db 958 GGAATTAGTGTGCTCAATTCAAAGTCGCAATATCTTAAATAC--ATTATCTAC 1014
QY 109 ThrGlyGlnTPraLeuThrLeuTPraSerGlyAspLeuValArgGlnArgSerIle 128
Db 1015 ATTGCTCAAAACCTTGTACTACAGCTTCA-----GCTTCTTCAAAATTC 1062
QY 129 SerGlyValAsnThrAlaHisThrProSerProValAlaGlnSerSerArgPro 148
Db 1063 GGTGCTTCAACCAATTC-----GCAGACACTACCAACC 1098
QY 149 ValGlnGlnHisProAlaValGlnGlyProThrProProValAlaVallyLeuPro 168
Db 1099 ACTTCTGTGACACCTGCA-----AAACCACTTCACAAACACTGTTAAAGTTAAATCC 1152
QY 169 ThrProThr-----ProProValGlnGlnProAlaProAlaProProValThr 186
Db 1153 GGAATACCTTTGGGGCGCTTACGATTAATAATAACTAGTATGCTCAATGAAAGT 1212
QY 187 GluAlaProPheAlaThrGlySerSerGlyValMetGln-----PheArgTyPro 203
Db 1213 TGAATCATTTAAGTACGATACATTATATGTGTAATCTTATGTTCAATCT 1272
QY 204 ValGlyAlaThrAsnProValAlaArgArgPheGlyThrAlaThrValAlaGlySer 223
Db 1273 GCTGCTGCTTCAAAATCTC-----TCGACAGGTTCCAGGCTCAACT 1311
QY 224 ValThrSerAsnGlyMetTPraPheSerGlyArgAspGlyAspLeuIleAsnAlaSer 243
Db 1312 GCTACCAATAC-----TCAAAC 1329
QY 244 AlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGln 263
Db 1330 TCG----- 1332
QY 264 ThrAsnGlyPheValSerSerTyThrIleHisIleLeuAspAlaGlnVallyThrGly 283
Db 1333 ACTTCTTCAATCAAAATGCTCAATTCAT-----AAGTGCTTAAAGGAGAT 1380

```

```

QY 284 ThrVal-----ArgThrGlyGlnArgIleAlaSerMetLys 295
Db 1381 ACTCTCGGGAGACTTTCGCAAAATCTGCGACCGCAATTCCTTCAATCAACAG 1431

RESULT 6
US-08-737-716-11
Sequence 11, Application US/08737716
Patent No. 5955258
GENERAL INFORMATION:
APPLICANT: GILBE BUIST
APPLICANT: Gerard VENEMA
APPLICANT: Jan KOK
TITLE OF INVENTION: Adriamus Marinus LEDERBER
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737, 716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/NL95/00170
FILING DATE: 12-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94201353.3
FILING DATE: 12-MAY-1994
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
STRAIN: MG1363
IMMEDIATE SOURCE:
CLONE: Fig.4 - ORF8
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1823..1928)
US-08-737-716-11

Alignment Scores:
Pred. No.: 0.00412 Length: 1930
Score: 130.50 Matches: 73
Percent Similarity: 37.71% Conservative: 39
Best Local Similarity: 24.58% Mismatches: 96
Query Match: 7.81% Indels: 89
DB: Gaps: 14

US-10-018-706-2 (1-322) x US-08-737-716-11 (1-1930)
QY 26 ThrThrcysileuallaglycysalaserlyserprothryraenserthr----- 42
Db 727 ACTGCTGCTTTGACGGGAGATATGCTACCGACCAAGTATGCGCTTCACTGAATCGC 786
QY 43 -----SerGlySerGlySerHisArgThrSerGly 52
Db 787 ATTAATTTCTCAATATAATTGACTCGTTTGACGAGCTTCTTCACTGGAATATCTAAT 846

```

```

QY      53 SerGlyValLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAm--- 71
      847 TGTGGTGGC-----TCGACACCAACAATTACAAATTAATTTCTGGACCAATAC 897
QY      72 -----ArgTyrGlnValIleGlnGlyAspThrValSerIleAlaGlnArgTyr 88
      898 ACTTCACTACTATACCGTCAATCTGTGATACCTTTGGGGAATGCACAAAGATAT 957
QY      89 GlyLeuAsnTrpArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyr 108
      958 GGAATATAGTGTGCTCAAAATTCAAAGTCGAAATTAATCTTAAAGTACCC--ATTATCTGC 1014
QY      109 ThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeuIleValArgGluArgSerIleSer 128
      1015 ATTTGGTCAAAACTTGTACTGACAGGTCA-----GCTTCTTCTACAAATTC 1062
QY      129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
      1063 GGTGGTTCAAACCAATTC-----GCAAGCACTACTCCAAACC 1098
QY      149 ValGlnGlnHisProAlaValGlnIleProThrProProValValValIleValysPro 168
      1099 ACTTGTGTGACACCTGCA-----AAACCACTTCACAAACAAGCTTAAAGTTAAATCC 1152
QY      169 ThrProThr-----ProProValValGlnGlnProAlaProValAlaProProValThr 186
      1153 GGAAGATACCTTGGGCGCTATCAGTAAATATAAAGTAACTAGTATCTCAATGGAAAGT 1212
QY      187 GluAlaProPheAlaThrGlySerSerGlyValMetGln-----PheArgTyrPro 203
      1213 TGGATCATTTATGTTCAATACCATTTATATGTTCAAAATCTATTGTTTCAACATCT 1272
QY      204 ValGlyAlaThrAsnProValValAlaArgArgPheGlyThrAlaThrValAlaGlySerThr 223
      1273 GCTGCTGCTCAAAATCT-----TCGACAGGTTCAAGCTCAACT 1311
QY      224 ValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
      1312 GGTACCAATTAAC-----TCMAAC 1329
QY      244 AlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis 263
      1330 TCG----- 1332
QY      264 ThrAsnGlyPheValSerSerTyrIleHisIleIleValAspAlaGlnValIleThrGlyAsp 283
      1333 ACTTCTTAACCTCAATATGCTCAATTCAT-----AAGTCTTAAGAGAT 1380
QY      284 ThrVal-----ArgThrGlyGlnArgIleAlaSerMetIle 295
      1381 ACTCTCTGGGGACTTCGCAAAATCTGGCAGCCCAATGCTTCATCAATCAAG 1431

```

RESULT 7
US-09-105-537-1/C

Sequence 1, Application US/09105537A

Patent No. 6255202

GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600,438US1

CURRENT APPLICATION NUMBER: US/09/105,537A

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 15872

TYPE: DNA

ORGANISM: Streptomyces venezuelae

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
34	130.50	38.75%	25.31%	7.81%	4	15872	81	43	115	81	18

US-10-018-706-2 (1-322) x US-09-105-537-1 (1-15872)

```

QY      34 AlaSerIleProThrTyrAsnSerThrSerGlySerHisArgThrSerGlySer 53
      3195 GCGTGGCGCCC-----TCGGGACAGACCGCCGACAGGCTCTCGGGGGAG 3151
QY      54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
      3150 GTGACCGCGCGGGTAGCGGACGAGCCATGCCAGATATCGATGCGATGCGCGGTGTG 3091
QY      64 ThrAsp-----SerGlnGlyValProAsnArgTyrGlnValIleGlnGlyAspThrVal 81
      3090 TCCGCGCGGGGTGCGGGGGGTATCCCTGTGCGCCATCCGATCTCGCGCGCGCGGTG 3031
QY      82 SerIle-----IleAlaGlnArgTyrGlyLeuAsnTrpArgGlu----- 94
      3030 ACCAGGTGCGCCAGGTGGCGCGGACGCGCGCGCGCGCGGTGTGCAAGACAGTCCG 2971
QY      95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThr 114
      2970 CTGGGACAGCGGACGCTCGGTGTGTGACAGACCGCGTCCGACG----- 2926
QY      115 LeuTrpSerGlyAsp-----LeuValAlaArgGlnArgSer 126
      2925 -----TCGACGCAATGAGAGATCGAAGCCGACCTCTGAAGTGAAGCCGAGTTCC 2872
QY      127 IleSerSerGly-----ValAsnThrAlaHisThrProSerPro 139
      2871 ACGCGGTGCGGTCTGCTCTACTGACGAGCGCGGCGAGTGTGCGCGACCGCTCGGTC 2812
QY      140 ValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValAlaGlnIleProThr 159
      2811 GTCTGTGCGCGGCGCGCGCGATCACCG-----CCACCGCGACCGCGCGCTCCG 2764
QY      160 ProProValValValIleValIleValIleProThrProThrProProValValGlnIleProAla 179
      2763 CCC-----TCGCGCTCGCGCTGCGCGCGATGACACCGCGCGCGG 2722
QY      180 ProValAlaProProValThrGlnAlaProPheAlaThrGlySer----- 194
      2721 CCGGTGCGCGGACCGGTCCCGATCGCCGCCGTGACGCGCGCGCGCTTCCGCGCGG 2662
QY      195 -----SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProValVal 211
      2661 CCGTCAACCAAGTGGCGTTCGCGCTGGAAGCGTAAAGTGGGAGGAGCAACCTGACCTG 2602
QY      212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
      2601 CCGGTGCTCCGTGCGCGCGGTCCATGTCAGTGTGCGCGCGCGGACGAAAGAGGTGTG 2542
QY      226 Ser-----AsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
      2541 AGTGGGCGGACGAGCATGGCTGCGCGCGGCGG-----TTGCGCAGG 2497
QY      244 AlaGlyThrValIleGlnAlaAsp-----HisAsnMetAspGlyAlaSerIleValIleGln 262
      2436 GCGGAGACCGCGGTGCGCGCTCTGCTGTGCGGTACGAGTCCGCGCATGCGCGGAGCG 2437
QY      263 HisThrAsnGlyPheValSerSerTyrIleHisIleValIleAsp-----AlaGlnValIleThr 281
      2436 ACCCGTGGGACGAGCTCCAGAGAGGTGTGCGCGCGCATTCCTCCAGGAGGTGTGAG 2377
QY      282 GlyAspThrValArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyAla 301
      2377 ----- 301

```


Db 2376 GCGTCAGGAAGCGTACGCGCTCGCG-----ACCTGTCACCACTGCGGATCG 2323

RESULT 8

US-09-103-840A-2/C

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Alignment Scores:

Pred. No.:	654	Length:	4403765
Score:	130.50	Matches:	81
Percent Similarity:	34.37%	Conservative:	41
Best Local Similarity:	22.82%	Mismatches:	128
Query Match:	7.81%	Indels:	106
DB:	4	Gaps:	17

US-10-018-706-2 (1-322) x US-09-103-840A-2 (1-4403765)

Qy 40 AenSerThrSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGly 59

Db 431307 AACCTGGGACCGGACCTGGGATCGGCAACTCGGACGCGGACACCGGCACTCGG 431248

Qy 60 SerGlnValIleThrAsp-----SerGlnValProAsnArgTyrGlnVal 75

Db 431247 AACACCGGACGACCTAACACCGGCTTCTTCAACACCGGACCTGTAACACCGGCTGCC 431188

Qy 76 LysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTyrArgGlnIle 95

Db 431187 AACGGGGGACG-----TACAACACCGGCTGGTGAACAACAC 431152

Qy 96 GlyHisIleAsn----- 99

Db 431151 GCGACACCAACACCGGACCTGGGCACTTCAACACCGGCTTCAACAACAC 431092

Qy 100 --AenLeuAsnSerGlyThr-----IleTyrThrGlyGlnTyrLeuThr 114

Db 431091 GGCATTTCAGTACGCGGCTTCCCAACAGGAGTATTCGCCACCGGCGCTTATCACC 431032

Qy 115 -----LeuTyrSerGlyAspLeuLysValArgGlnArgSerIle 127

Db 431031 GCGGACATGGGCAACGCGGCTTTCGCGGCGGAC-----CACACGCGCTTATTC 430981

Qy 128 SerSerGlyValAsnThr-----AlaHisThrProSerProValAla 141

Db 430980 AGCGGGGCTATCGGGCTCATGTTCCGAATAACCGGACACGTCACCGGTGAAGTTCC 430921

Qy 142 ValGln-SerSerArgProProValGlnGlnHisProAlaValGlnLysProThrProPr 161

Db 430920 GTCAACATCCCATACCGGACGCT-----TACCAACACCGCTC 430882

Qy 161 ovalValVal-----ValLysLysProThrProThrProProValGlnProAl 179

Db 430881 TACAGCGGACATACCGCTTGAGCAATCAACTCGGTTTCAACATGACATCGACAGGATC 430822

Qy 179 aProValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetG 199

Db 430821 CCCCTGCTGGGCTGATGATGACGAGCCGCTTCCGCCCATACCGGACCGGTCC 430762

Qy 199 nPheArgTyrProValGlyAlaThr-----AsnProValAlaArgArgPheGlyTh 216

Db 430761 GCGATACCGGTCAACATCGGAGCACCTGGCGGCTTCAACCGG--ATCAGGATCCGGGCAC 430703

Qy 216 rAlaThrValAlaIleSerThrValThr----- 225

Db 430702 CGAAGGCTGCTGCTTCGATGTCACGTTTCGTCACATTCGCGCTACACGCGCTTTT 430643

Qy 226 -----SerAsnGlyMetTyrPheSerGlyValArgAspGlyAspLeuI 239

Db 430642 CAAGCCACCAACCGATCCGCTCTCGGC-----TTCTTCAACGCGCGCCCGGACCGTATC 430586

Qy 239 eAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerI 259

Db 430585 GGGCATCGCCACATCGCGGCAACATTCGCGGCTTCCAGAACGTCGGAACCTCGCG-- 430528

Qy 259 eValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnVa 279

Db 430527 -----ACCTCGGCTTC--AACACTACGCGCTCGCTGCAATCGGACCTGCG 430484

Qy 279 lLysThrGlyAspThrVal-----ArgThrGly-----GlnArgG 291

Db 430483 GAACCTGGGCGATACCGCTTCGCGGCTATTCACACCGGACCTCGGGGACACCGGCAACGT 430424

Qy 291 eAlaSerMetLysAsn-----GlnProSerG 300

Db 430423 CTCGGGACATGTTCAACATCGGACGACCACTCGGCGGCTTCTTCCACGACGACGAGGACCG 430364

Qy 300 yAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyr 314

Db 430363 GATGTCATGTTCAACTCGGCTCGGCAACATCGGCAATTC 430321

RESULT 9

US-09-103-840A-1/C

Sequence 1, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 1

LENGTH: 4411529

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37Rv

US-09-103-840A-1

Alignment Scores:

Pred. No.:	656	Length:	4411529
Score:	130.50 <td>Matches:</td> <td>81</td>	Matches:	81
Percent Similarity:	34.37% <td>Conservative:</td> <td>41</td>	Conservative:	41
Best Local Similarity:	22.82% <td>Mismatches:</td> <td>128</td>	Mismatches:	128
Query Match:	7.81% <td>Indels:</td> <td>106</td>	Indels:	106
DB:	4 <td>Gaps:</td> <td>17</td>	Gaps:	17

US-10-018-706-2 (1-322) x US-09-103-840A-1 (1-4411529)

Qy 40 AenSerThrSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGly 59

Db 431224 AACCTGGGACCGGACCTGGGATCGGCAACTCGGACGCGGACACCGGCACTCGG 431165

QY	60	SeGIhVal1lleThrasp-----SeGIhGlyVal1ProAsnIhgrVal1	75
Db	431164	AACACCGGAGCACTAACCGGGCTTTTCAACACCGGATCGTCAACCGGGTGTCCG	431105
QY	76	LySGhGlyYAspIhValSerIyS1leAglhIhgrIyGlyeuhentIhgrGlu1le	95
Db	431104	AACGGGGGAGC-----TCAACACCGGCTGTGTCAACAC	431069
QY	96	GlyIhS1leAsn-----	99
Db	431068	GGGACACCAKACCGGATCGGCAACTGGGCACTTCAACACGGGCTTCAACAC	431009
QY	100	---AspIeuhanserIyThr-----IleTyThrGlyhIhTpleuThr	114
Db	431008	GGCAATTTCAGTACGGGGCTTGTCCAACACGGTATGTATGCCACCGGGCTTTTCAAC	430949
QY	115	-----LeuTPSerGlyYAspLeuLyvalAArgIuhArserIle	127
Db	430948	GGCGACATGGGACACCGGCGCTTCTGGCGCGGAC-----CAGACGGCGCTATT	430898
QY	128	SerSerGlyYValAsnThr-----AlahIshIhProSerProValAla	141
Db	430897	AACGGGGGTATTCGGGTCCATGTTCCGAATACCCGGCACCGTCAACCTGGAAATTC	430838
QY	142	ValGln-SerSerArProProValGlnGlnIhIhSproAlaValGlnLyProThzPro	161
Db	430837	GTCAACATCCCATCATCCCGGACCT-----TCAACCAACACCGTC	430799
QY	161	oValValVal-----ValIySerProThzProThzProProValValGlnIhProAl	179
Db	430798	TACAGCGGATACACGCTTAGGAAATCAACTTCGGTTCAACATGCACATCGAGGATC	430739
QY	179	ArProValAlaProProValThhGlnIhAphroAlaIhThhGlySerSerGlyValMetel	199
Db	430738	CCCTGTCTGGCCGGTGCATACGACAGGCGCTTCTCCGCCCATCACCGGACCGGTCCC	430679
QY	199	nPherArGlyrProValGlyValIaThr-----AsnProValValAArgArPhheGlyTh	216
Db	430678	GGCATACGGTCAACATCGGCGACCTGGCGGTTGCACGCG-ATCAGATCCCGGCAC	430620
QY	216	rAlaThrValAlaGlySerThrValThr-----	225
Db	430619	CGCAAGCGTGGTCCCTTGATGTACAGTGTCAACATTTGGGCTACACACGGGCTTTT	430560
QY	226	-----SerAsnGlyMetIhThzPheserGlyArAspGlyYAspLeu1	239
Db	430559	CACGCCACCAACCGATCCGCTCTCGGGC---ATTCTTCAACGGCGGCCGCCGAAACGTATC	430503
QY	239	eAsnAlaSerAsnAlaGlyThrVal11eGlnIhAspIhAsnMetAspGlyYAlaser1	259
Db	430502	GGGATCCGCAACATCGGCGGCACAACTTCCGGCTTCCAGAACGTGCGCACTCCGCG-	430445
QY	259	eValIleGlnIhIhThrasnGlyPhValaserIyTylIehIhS1leYAspAlaGlnVa	279
Db	430444	-----ACCTGGGCTTC---AACAACTACGGCTGTGTCAATCGGAGCTGGC	430401
QY	279	IllyThrGlyAspThrVal-----ArghhGly1	291
Db	430400	GAACTGGGCGATACCGTCTGGGGGTATTCAACACCGGATCGGGCACCGGCAACGT	430341
QY	291	eAlaserIyAsn-----GlnProSerG1	300
Db	430340	CTCGGCACTTTCAACATCGGACGACAACTCGCGGGTTTTCACGACGACGGACCGG	430281
QY	300	YAlaAlaIeuhheGluIhPherGlyIaserArghsnGlyValIyTr	314
Db	430280	GATGCGATGTTCACCTCGGCTGTGGAAACATCGGCCCAATTC	430238

RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328

```

/ GENERAL INFORMATION:
? APPLICANT: FLEISCHMAN, Robert D.
? APPLICANT: WHITE, Owen R.
? APPLICANT: FRASER, Claire M.
? APPLICANT: VENTER, John C.
? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
? TITLE OF INVENTION: TUBERCULOSIS
? FILE REFERENCE: 2436-2007.00
? CURRENT APPLICATION NUMBER: US/09/103.840A
? CURRENT FILING DATE: 1998-06-24
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 4411529
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37RV
/ US-09-103-840A-1

Alignment Scores:
Pred. No.: 3.11e+03
Score: 123.50
Percent Similarity: 34.21%
Best Local Similarity: 27.89%
Query Match: 7.40%

Length: 4411529
Matches: 53
Conservative: 12
Mismatch: 74
Indels: 52
Gaps: 10

```

US-10-018-706-2 (1-322) X US-09-103-840A-1 (1-4411529)

OY	135	HieHrProSePrOvAlAlaValGlnSerSerXgPrOvAlGlnGlnHisProAla	154
Db	190636	CACCCGCCAGAACCA---GCCGCCCGATGAGCAGCCGGGTCCCGCGCGCCG---	1990689
OY	155	VaIGlnLysProThrPrOvAlVaIValVallLysProThr---ProThPrOPro	173
Db	190690	-----CCGACGCCGCCTGTGTGTGTGGCGCCCCCGCACCAACCGTACC	1990740
OY	174	-----ValValGlnGlnProAlaProValAaProprovalThrGluAlaPro	189
Db	1990741	GAACCATCAAAGAAGCGGATTCGCGCGGCGCACCGCGCCCGACCGCGTTCCG	1990800
OY	190	PheAlaThrGlYserSerGlYalMetGlnPheArgTYr-----ProValGly	205
Db	1990801	AC-GCCCGCGAGCCCGCGCGCGCCACCGGTGCCAAAGAGATCCCGCTGCGCCACCGCG	1990859
OY	206	Ala-----ThraSProValValAArgArgPheGlyThr	216
Db	1990860	GCCGCCCGCGCTGCGCGTGTCGCCGCGCGCACCGCGCCCGCGTTGCCGCGTGGCCGAA	1990919
OY	217	AlaThrValAlaGlYserThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGly	236
Db	1990920	CAACCAAGCCCGCGCACCGCGCATC-----	1990943
OY	237	AspleuLeaSnAlaSerAsenAlaGlyThrValIlleGlnAlaSPHisaenMetAspGly	256
Db	1990944	----::gtccccggattcccgccgcgccacatctgcgcgttacccagtacggcgggctcggt	1990997
OY	257	AlaSerIlleValIlleglnHisThrAnGlyPheValSerer-----Tyrlle	272
Db	1990998	-----CAATCCCTCCGGTGGGTTCTGTGATGATAACTGAGATGTCTGCTG	1991042
OY	273	HislelLeYaSpAlaGlnValLyThrnglyYasPthValAArgThrglyGlnaArglleAla	292
Db	1991043	CAGATTGTGCATGGCAGGTGCTCTCGGAGAGCGTTATATCCGTGGC-----	1991090
OY	293	SerMetLysAnGlnProSerGlyAlaAla	302
Db	1991091	--GCCACGACGACAACCCCGCGAAGCGCC	1991117
RESULT	11		
US-08-276-213-6			
Sequence 6,			
Patent No. 5536655			
Accretion us/08276213			

RESULT 11
US-08-276-213-6
; Sequence 6, Application US/08276213
; Patent No. 5536655

GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Himmell, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-6

Alignment Scores:
Pred. No.: 0.0755 Length: 3004
Score: 120.50 Matches: 96
Percent Similarity: 33.98% Conservative: 27
Best Local Similarity: 26.52% Mismatches: 118
Query Match: 7.22% Indels: 122
Gaps: 20

US-10-018-706-2 (1-322) x US-08-276-213-6 (1-3004)

Qy 27 Thrcyslleleuajaglycyaleaserlyprothrtysraserthserglysergly 46
Db 1629 ACGTCCGACACCGCTGCTGT-----ACTCCGCGACGACGACGACGACG 1676
Qy 47 Serhisargthnserglyserglyleualalleglyserglnvalillthrasper 66
Db 1677 TCTACCCGACGCTGCTCA-----GCGATCCGACCTTCC 1712
Qy 67 Glnglyvalproasnargtyrclnvallysglnglyaspthrvalser---lyslela 85
Db 1713 CCAACAACATGCGCGCATCTGGAACAGACGATCGGATACCTCTTCATCATGAAATG 1772
Qy 86 Glnargtyglyleuasntpr---argglulleglyhislelaen----- 99
Db 1773 CACCGTATGCG-----TGCGCAATTCGATACGACACTGCAATCCAGACGACG 1826
Qy 100 -----Asnleuasntserlyr-----Thrltlythrly 110
Db 1827 CGTGCCTGAACGCTCTCCATGATCTAATGCGGACGCGCAATTCGCTGCGACACT 1886
Qy 111 --Glntrpleuthrltlyrser----- 117

Db 1887 TCCAGTG3--ACCTCTGTCTGGAACCCCGATCCGCGACAGAGAAATTTCA 1943
Qy 118 -----Glyaspyleuvalarggluargx 126
Db 1944 AGATGACTGCAGACGCTGCACAGTAAGACGGCTATCTGCCCGATCAAGTGT 2003
Qy 126 erlle-----serglyvala 132
Db 2004 CGATTTTCGATCCTGTGCGCGCTGTGATCGCTGACGATCAACCGTCCCGTGGT 2063
Qy 132 snthralahsrthproserprovalalavalglnserargproprovalglnh 152
Db 2064 CGCCGCTCCGCTCCGACGCGCTGCGG-----AGTCGACGCGACGCTTACTC 2114
Qy 152 isproalavalglnlybprothproprovalvalvallyblybprothpro 172
Db 2115 CGACGCGACAGCCAGCCCGACGCA-----ACGTCGACCTTACTGCTACGCGC 2168
Qy 172 ropovalvalglnlybproalaprovalala-proprovalthrualapro 191
Db 2169 CCAAGGACAGCCGACGCGCTGCACGACGAGCTCCGAGCCCGCTGACCGGAGT 2228
Qy 192 Thrlyserse-----glyvalmetglnpheargtyrprovalglyalathrasnpro 209
Db 2229 ACCAGTCAACAGCGATTTGGGGCAATGCG-TTCAAGGTACGATGCGCGTCAAAAT--- 2284
Qy 210 valvalargargpneglythralathrvalalaglyserthvalthrseranglymet 229
Db 2285 -----TCCGATCCGTCGCGACCAAGACATGACGCTGAT--- 2320
Qy 230 Trptheserglyargaspglyaspyleuileasnalaaserenalaglythrvalillgln 249
Db 2321 TGACATTCGCGGGAATCAGACATTAACATTCGTCGATGACGCGTC----- 2371
Qy 250 Alaasphisasnetaspglyalaserlle-----Val 260
Db 2372 -----ACGCAAGACGCTGACGCTGATACGCGCTGCAATGATGATTAACAACGTCG 2422
Qy 261 lilegln-----Histhranglyphealaserlyrille----- 272
Db 2423 ATTCAGCTGCTGCAACACACGCTCGGATTCAGAGCGACGACTATACCGGACGACGCG 2482
Qy 273 -----Hislelyasapala-glnvallythrlyaspthrvalargthrlygln-- 289
Db 2483 GCACCGACAGTGGCTGCGACGACGATTAACCTCGGGAGCCGACGAGGCTCCGGA 2542
Qy 290 -----Argillealasermetylaspnglnproserglyalala 302
Db 2543 CCGTCGCTCCCGGCTTCCACTATGAGGCAACCAACATCCGACGGAACGCA 2600

RESULT 12
US-09-136-574A-1
Sequence 1, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESS: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

Qy 237 Aspleuileasnaalaserasnaalaglythrvalileginalaasphiasmetaspoly 256
Db 1988737 -----GTCCCGGTTCCCGCGGCCCACTCTCCCGCTTACCATCAGCGGCGCTCGGT 1988790
Qy 257 Alaserilevalileginalaasphiasmetaspoly 272
Db 1988791 -----CATGCGCTCGGTGTTCTGTTGTAAGTAAGTAAGTAAGTCTGCTG 1988835
Qy 273 Hieileysasphialaginalaasphiasmetaspoly 292
Db 1988836 CAGGTGTCGATGCGGAGGTCTCTCGGAGCGCTTATATCCGTGCGC----- 1988883
Qy 293 Sermetlysaenglinproserglyalaala 302
Db 1988884 ----GCCAGCAGCAACCCCGGAGCGCC 1988910
RESULT 14
US-09-462-606-1
; Sequence 1, Application US/09462606
; Patent No. 6432408
; GENERAL INFORMATION:
; APPLICANT: MENG, XIANG-JIN
; APPLICANT: Emerson, Suzanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: A SWINE HEPATITIS E VIRUS AND USES THEREOF
; FILE REFERENCE: 20264267US1
; CURRENT APPLICATION NUMBER: US/09/462,606
; CURRENT FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: US 60/053069
; PRIOR FILING DATE: 1997-07-18
; PRIOR APPLICATION NUMBER: PCT/US98/14665
; PRIOR FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 5127
; TYPE: DNA
; ORGANISM: Hepatitis E virus
US-09-462-606-1
Alignment Scores:
Pred. No.: 0.301 Length: 5127
Score: 118.00 Matches: 79
Percent Similarity: 31.54% Conservative: 44
Best Local Similarity: 20.26% Mismatches: 129
Query Match: 7.07% Gaps: 138
DB: 4 Gaps: 17
US-10-018-706-2 (1-322) x US-09-462-606-1 (1-5127)
Qy 1 Methrvalthrlealeasnaenglinasphiasmetaspoly 20
Db 1642 CTACCGGCTGCTGAGCTGCTGAGCAAGCCAGACCGCTTATAGTGCAGCTGCTT 1701
Qy 21 Iileheglyvalilethrthrlealeasnaenglinasphiasmetaspoly 37
Db 1702 GGTATATAGACCTTACAGACCAAGCTGTATAGTGCAGCTTATAGTGCAGCTTAT 1761
Qy 38 Thrtyranserthrserglyserglyserglyserglyserglyserglysergly 57
Db 1762 GAGCAGTATGCTATCATCTGAGCGCTCCGCGCTGATGAGGCGGCGGCTGATGAG 1821
Qy 58 Ile-----Glyserglyvalilethrthrlealeasnaenglinasphiasmetaspoly 70
Db 1822 CTCACTATGAGCTACCCCTGCGGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1881
Qy 71 Asnarglyrvalinvalynglyasphiasmetaspoly 82
Db 1882 TGACAGCAGCATTCGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1941
Qy 83 -----Lysilealeaglnarglyrvalyngly 91
Db 1942 TTTTGACGTGCTTAT 2001

Qy 92 ---Tprarg-----Gluileglyhile 98
Db 2002 TTATGCTACACCTGAGGATTTCTGGGATCTTCCCGCTTCTCCCGGCGGCACTT 2061
Qy 99 -----Asnaenleuasnaenglinasphiasmetaspoly 114
Db 2062 TGGAGCGCTGTAACCTTTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2121
Qy 115 Leutpserglyaspleuysvalarglnargserileserglyvalasnaethala 134
Db 2122 -----TCGGCTTTCTAGGAT 2139
Qy 135 Hiehrproserprovalalavalalnsersearproprovalalnglnhileproala 154
Db 2140 TTCTCCCGCT-----GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2187
Qy 155 Valglnlyspthrproproval----- 162
Db 2188 CTGCGCCACCTTACCCCGCTTATGATGATGATGATGATGATGATGATGATGATGAT 2247
Qy 162 ----- 162
Db 2248 TCTCAGGTGATGCGGATCTGTCCTCCCTGTCGAGCCGCTGATTAACCGAGCTCC 2307
Qy 163 Valvalalylslyspthrprothrprothrprothrprothrprothrprothrprothr 182
Db 2308 ATTGTGCTTACCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2361
Qy 183 ProprovalthrvalalprophalaThrglyserglyvalmetglnphearglyr 202
Db 2362 CCGGCT-----TCCGCGAGCTGTGTCTCTCTACACCTAT 2397
Qy 203 Provalalylalathrasnaenprovalvalarghrpheglythrvalalalaglyser 222
Db 2398 CCCGCGCGCGCA-----AAGGTGTATGCGGCGGCTCA 2427
Qy 223 Thrvalthrserasnaenglymettrpserglyargasphiaspleuileasnaaser 242
Db 2428 TTGTTTATATCATCATCTTATCTG-----CTGTTATATCTCA 2466
Qy 243 Asnaalaglythrvalileginalaasphiasmetaspolyalaserilevalilegln 262
Db 2467 AACCGGCG-----CACCGCGCTGAGGAGTGC-----CTCTGC 2499
Qy 263 Hiehrasnglyphevalasersearlyrilehileysasphialaglnvalythrgly 282
Db 2500 CAT-----GCTTTTACCAAGTTTCCAGAGGCGGTTTACCCGAGCTGAG 2544
Qy 283 Asphthrvalargthrvalarglnalalasermetlysaenglinpro-----serglyala 301
Db 2545 TTCACTATGCTGAGGCGCTTGCAGCATATATATATATATATATATATATATATAT 2604
Qy 302 Alaleuphegluphearglyserargasn 311
Db 2605 GTGGCGCGCGCATATATAGGTTGAGCAGAT 2654
RESULT 15
US-09-462-606-3
; Sequence 3, Application US/09462606
; Patent No. 6432408
; GENERAL INFORMATION:
; APPLICANT: MENG, XIANG-JIN
; APPLICANT: Emerson, Suzanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: A SWINE HEPATITIS E VIRUS AND USES THEREOF
; FILE REFERENCE: 20264267US1
; CURRENT APPLICATION NUMBER: US/09/462,606
; CURRENT FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: US 60/053069
; PRIOR FILING DATE: 1997-07-18
; PRIOR APPLICATION NUMBER: PCT/US98/14665
; PRIOR FILING DATE: 1998-07-17

NUMBER OF SEQ ID NOS: 65
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO: 3
 LENGTH: 7207
 TYPE: DNA
 ORGANISM: Hepatitis E virus
 US-09-462-606-3

Alignment Scores:
 Pred. No.: 0.511 Length: 7207
 Score: 118.00 Matches: 79
 Percent Similarity: 31.54% Conservative: 44
 Best Local Similarity: 20.26% Mismatches: 129
 Query Match: 7.07% Indels: 138
 Gaps: 17

US-10-018-706-2 (1-322) x US-09-462-606-3 (1-7207)

```

QY 1 MetThrValThrIleAlaIleAsnSerGlnAsnGlnIleProIleLysArgLeuGlyLeu 20
DB 1651 CTAAACGGCTACTGTTGAGCTCACTGCAAGCCGAGCCGTTTAGAGTCCGCGACTGCTT 1710
QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCys-----AlaSerLysPro 37
DB 1711 GGTAATPAGACCTTCAGACGACGGGTGATGGCGCCCATCTTGAGCGCAATGTCTCT 1770
QY 38 ThrTYAsnSerThrSerGlySerHisArgThrSerGlySerGlyLeuAla 57
DB 1771 GAGCACTATGCTCTATCATTCACAGCGCTCCGCCCATCTATGAGGCGCGGTCACTAC 1830
QY 58 Ile-----GlySerGlnValIleThrAspSerGlnIleValPro 70
DB 1831 CTCACCTATGAGCTACCCCTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1890
QY 71 AsnArgTYGlnValIleGlnGlyAspThrValSer----- 82
DB 1891 TGCACAGCCACATCCCGCGCGGCGCGCTAGGCGTGGCGCGGAGGAGGAGGAGGAGG 1950
QY 83 -----LysIleAlaGlnArgTYGlyLeuAsn----- 91
DB 1951 TTTTGAGTGCCTTATATAGATATATATAGTTCACACCGGCACTGCGCGGACAT 2010
QY 92 ---TTPArg-----GluIleGlyHisIle 98
DB 2011 TTATGGCTACACCTGAGGAGATTTGCGGCACTTCCCTTTCCTCCCGGACAT 2070
QY 99 -----AsnAsnLeuAsnSerSerTYRThrIleTYRThrGlyGlnThrLeuThr 114
DB 2071 TGGAGACCTGCTAACCCCTTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2130
QY 115 LeuTPSerGlyAspLeuLysValArgGlnArgSerIleSerSerGlyValAsnThrAla 134
DB 2131 -----TCTGGCTTTTCTAGCAT 2148
QY 135 HisThrProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAla 154
DB 2149 TTTCTCCCCCT-----GAGGCGCGCGCGCGCGCTTTGGCGCGTCCCGCGGG 2196
QY 155 ValGlnLysProThrProProVal----- 162
DB 2197 CTGCCCCACCTTACCCCACTCTGTAGTACATTTGGGTGTACACACACTTCATAAGAG 2256
QY 162 ----- 162
DB 2257 TCTGAGTGCATGCGGACATGTGCCCCCTGCTCTGAGCCCGGTGATTACCGAGCTCC 2316
QY 163 ValValValLysLysProThrProThrProProValValGlnGlnProAlaProValAla 182
DB 2317 ATTGGCTTACCTCCCGCGCGCGCGCTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGG 2370
QY 183 ProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTYR 202
DB 2371 CCGGCT-----TCCCGCACTCGTGTCTCTCTCTACACCTAT 2406

```

```

QY 203 ProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySer 222
DB 2407 CCCGACGGCGCA-----AAGGTGATGCGGGGTCA 2436
QY 223 ThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSer 242
DB 2437 TTGTTGAATCAGACTGTACTGCTG-----CTGGTTAATGCTTCA 2475
QY 243 AsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGln 262
DB 2476 AACCGGGCG-----CACCGCGCTGGAGGTGGC-----CTCTGC 2508
QY 263 HisThrAsnGlyPheValSerSerTYRThrIleHisIleLysAspAlaGlnValIleThrGly 282
DB 2509 CAT-----GCTTTTACCAACGTTTCCGAGAGGCGTTTACCGAGCTGAG 2553
QY 283 AspThrValArgTYRThrGlyGlnArgIleAlaSerMetLysAsnGlnPro---SerGlyAla 301
DB 2554 TTCAATTAGCTGAGGCGCTTGCAGACATATACCTGACCCCGCCCTATCATTCATGCA 2613
QY 302 AlaLeuPheGlnPheArgIleSerArgAsn 311
DB 2614 GTGGCCCCCGCACTATGGGTGAGCAGAGAT 2643

RESULT 16
US-09-010-928B-1/C
Sequence 1, Application US/09010928B
Patent No. 5994099
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V
INVENTOR: Hayashi, Cheryll Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
ADDRESS: 8110 GATEHOUSE RD. SUITE 500B
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28997
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY:
LOCATION: 1..2830 /note= "Flagelliform DNA sequence
OTHER INFORMATION: taken from the 5' region. The putative start codon is at
OTHER INFORMATION: position 219"
FEATURE:
NAME/KEY: CDS
LOCATION: 219..2830
US-09-010-928B-1

```

Alignment Scores:
Pred. No.: 0.134 Length: 2830
Score: 117.50 Matches: 49
Percent Similarity: 36.04% Conservative: 22
Best Local Similarity: 24.87% Mismatches: 86
Query Match: 7.04% Indels: 40
DB: 2 Gaps: 7

US-10-018-706-2 (1-322) x US-09-010-928B-1 (1-2830)

QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerIlyProThyIraSerThySergIlySer 45
DB 2219 ACCGACCTGGTCCGAGGTCACACGACCTGACCGTAGGATCCACACGACACACC 2160
QY 46 GlySerHisArgThrSergIlySergIlyLeu-----AlaIleGlySergIlyValIle 63
DB 2159 GGGTCTCCGAAACCGGGTCTCCGAAACACGACGAGTCAAGTCTCCAGACGCGCT 2100
QY 64 ThrAspSergIlyValProAsnArgTyrgIlyValSgIlyAspThrValSerIly 83
DB 2099 TCCGCACTTCGGGTCGGAACACGACGAGTCAAGTCTCCAGACACGCGCTCC 2040
QY 84 IleAlaGlnArgTyrgIlyLeuAsnTrp-----ArgGluIleGlyHisIleAsn 99
DB 2039 TACGCTCCAGTCTTACGCGACCTGTCGAGACCACTAGTCTTACTCTCTGTC 1980
QY 100 AsnLeuAsnSergIlyThrIleTygThrGly----- 110
DB 1979 GGAGCCACGAGATCTTACGCTCTGTCGAGACCACTAGTCTTACTCTCTGTC 1920
QY 111 -----GlnTrp-----LeuThrLeuTrpSer 117
DB 1919 AAAGCTACAGAGTCTTACGCTCTGTCGAGACCACTAGTCTTACTCTCTGTC 1860
QY 118 GlyAspLeuIlyValArgGlnArgSergIlySergIlyValAsnThrAlaHisThr-Pr 137
DB 1859 -----AGAGCGACGAGTCAAGTCTTCTGTCGAGACCACTAGTCTTACTCTCTGTC 1821
QY 137 oSerProValAlaValGlnSergIlyProProValGlnGlnHisProAlaValGlnIly 157
DB 1820 GGGTCCGGAACCTCAGCGCTCTCCACCGATGTTAATCTTCTGATATGTTATTCGG 1761
QY 157 sProThrProProValAlaValIlySlySProThrPro-----ThrProProValAla 175
DB 1760 GCCATCTGCACATCAATGTAATATCAAAATCTCTACATGTTGTCGCGCGC----- 1706
QY 175 lGlnGlnProAlaProValAlaProProValThrGlnAlaProPhaIa 191
DB 1705 GATCTCTGCTGTCACACGACCTCTGCTCCACGACCACTCTAGCT 1658

RESULT 17
US-09-125-287-2
; Sequence 2, Application US/09125287B
; Patent No. 6114602
; GENERAL INFORMATION:
; APPLICANT: BARG, Rivka
; APPLICANT: SALTS, Tehnam
; TITLE OF INVENTION: METHOD FOR THE INTRODUCTION OF GENETIC PARTHENOCAPI IN
; FILE REFERENCE: INTRO GENETIC PARTHENOCAPI IN PLANTS
; CURRENT APPLICATION NUMBER: US/09/125.287B
; EARLIER FILING DATE: 1998-11-09
; EARLIER APPLICATION NUMBER: PCT/IL97/00051
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 4518
; TYPE: DNA
; ORGANISM: TPRP-F1 PROMOTOR
; FEATURE:
; NAME/KEY: unsure

LOCATION: (6)..(191)
; OTHER INFORMATION: "n"s are any nucleic residue
US-09-125-287-2

Alignment Scores:
Pred. No.: 0.387 Length: 4518
Score: 116.00 Matches: 35
Percent Similarity: 44.76% Conservative: 12
Best Local Similarity: 33.33% Mismatches: 40
Query Match: 6.95% Indels: 19
DB: 3 Gaps: 5

US-10-018-706-2 (1-322) x US-09-125-287-2 (1-4518)

QY 136 ThrProSerProValAlaValGlnSergIlyProProValGlnGlnHisProAlaVal 155
DB 3313 ACTCAACACCACTTATGTC-----CATTCACAGTCACTCCAAAACCA----- 3357
QY 156 GlnIlySProThrProProValAla-----ValIlySlySPro-----ThrProThr 171
DB 3358 CCATCAACCAACCTCTATGTTTACCCCATGTTTATCCAAATCACTCCAAACA 3417
QY 172 ProProValAlaGlnGlnProAlaProValAlaProProValThrGlnAlaProPhaIa 191
DB 3418 CCACTGTTGTCACCTCAATCATTCACCAACCACTATGTCCTCCACCTTTGTC 3477
QY 192 ThrGlySergIlyValMetGlnPheArgTyg----- 202
DB 3478 CCATCTCTCCGCTGTAATACCAACCACTTACCTGCAAGTCTCCGTTGTTATCCA 3537
QY 203 ProValGlyAlaThrAsnProProValAlaArgPheGlyThrAlaThrValAlaGlySer 222
DB 3538 CCATCACT-TCCAACACCCCTTACCATGCCACCAACCAACCAACCAACCAATAT 3596
QY 223 ThrValThrSerAsn 227
DB 3597 ACCATCAACCAACGAC 3611

RESULT 18
US-09-125-287-1
; Sequence 1, Application US/09125287B
; Patent No. 6114602
; GENERAL INFORMATION:
; APPLICANT: BARG, Rivka
; APPLICANT: SALTS, Tehnam
; TITLE OF INVENTION: METHOD FOR THE INTRODUCTION OF GENETIC PARTHENOCAPI IN
; FILE REFERENCE: INTRO GENETIC PARTHENOCAPI IN PLANTS
; CURRENT APPLICATION NUMBER: US/09/125.287B
; EARLIER FILING DATE: 1998-11-09
; EARLIER APPLICATION NUMBER: PCT/IL97/00051
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 12839
; TYPE: DNA
; ORGANISM: TPRP-F1 GENOMIC CLONE
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (5)..(11322)
; OTHER INFORMATION: "n"s are any nucleic residue
US-09-125-287-1

Alignment Scores:
Pred. No.: 1.95 Length: 12839
Score: 116.00 Matches: 35
Percent Similarity: 44.76% Conservative: 12
Best Local Similarity: 33.33% Mismatches: 40
Query Match: 6.95% Indels: 19
DB: 3 Gaps: 5

US-10-018-706-2 (1-322) x US-09-125-287-1 (1-12839)

QY 136 ThrProSerProValAlaValGlnSerArgProProValGlnGlnHisProAlaVal 155
 DB 3314 ACTCCACACACCTATTATGTC-----CATCCACGACGATCCCAAAACA----- 3358
 QY 156 GlnLysProThrProProValAla-----ValValLysLysPro---ThrProThr 171
 DB 3359 CCATCAACCAACACCTCTATTGTTTCACCCCATTTGTTATCCACCATCTCCACA 3418
 QY 172 ProProValValGlnGlnProAlaProValAlaProProValThrGlnLysLysPheAla 191
 DB 3419 CCACCTGTTGTCTACCTCCCAATTCATTCACACACACCTATTGTCCTCCACCTTTGTC 3478
 QY 192 ThrGlySerSerGlyValMetGlnPheArgTyr----- 202
 DB 3479 CCCAATCTCCCGTGTATATACCAACACCTACGTCGCAAGTCCTCCGGTGTACTCCA 3538
 QY 203 ProValGlyAlaThrAsnProValAlaValArgPheGlyThrAlaThrValAlaGlySer 222
 DB 3539 CCATAGT-TCCAAACACCCCTTACACCATGCCACACACACACACCAACCAATAT 3597
 QY 223 ThrValThrSerAsn 227
 DB 3598 ACCATCCACACACG 3612

RESULT 19

US-08-194-290-6
 ; Sequence 6, Application US/08194290
 ; Patent No. 5500353

GENERAL INFORMATION:

APPLICANT: Smtc, John
 APPLICANT: Bingle, Wade H
 TITLE OF INVENTION: Bacterial surface protein expression
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESS: Shlesinger, Arkwright & Garvey
 STREET: 3000 South Eads Street
 CITY: Arlington
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/194,290
 FILING DATE: 09-FEB-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Garvey, George A
 REGISTRATION NUMBER: 17737
 REFERENCE/DOCKET NUMBER: 5946-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-684-5288
 TELEFAX: 703-684-5600
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3300 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Caulobacter crescentus
 STRAIN: CB 15
 US-08-194-290-6

Alignment Scores:

Pred. No.: 0.297 Length: 3300
 Score: 115.00 Matches: 75
 Percent Similarity: 36.48% Conservative: 41

Best Local Similarity: 23.58% Mismatches: 134
 Query Match: 6.89% Indels: 69
 DB: 1 Gaps: 12

US-10-018-706-2 (1-322) x US-08-194-290-6 (1-3300)

QY 23 GlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsnSerThr 42
 DB 776 GGGCCGCTGTGACCGACCAACGCGCTGGCTGAACCTGTTCACCGCTATCCCTGCG 835
 QY 43 SerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnVal 62
 DB 836 GGGCGTGGGGTTCG-----ACCTCTCGCTGACACCGGACCGACACAC 880
 QY 63 IleThrAsnSerGlnGly-----ValProAsnArgTyrGln 74
 DB 881 CTGACGGGCGACCGCAACACGACGCTGCTGGGGGTGAAGTCGCGCGCGCTGCAC 940
 QY 75 ValLysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpArgGlu 94
 DB 941 CTGACCTGTGGCGACACCTGAGCGGGGCTGCTGGACCGACGCTGAACCTGGGTGCA 1000
 QY 95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThr 114
 DB 1001 GCTGCTGGGTTACGGCTCTCCGACCGCGTGAATC----- 1039
 QY 115 LeuTrpSerGlyAsnLeuLysValArgGluArgSerIleSerSerGlyValAsnThrAla 134
 DB 1040 -----TGGGCAATCAAAACGATG 1057
 QY 135 HisThrProSerProValAlaValGlnSer-SerArg-----Pr 147
 DB 1058 AACGTGACGTGGGGCGCTGCGCATCAACCTGAACAGCTTTCGGCGGTGACGGTCTGAC 1117
 QY 147 OProValGlnGlnHisProValAlaGln---LysProThrProProValValAlaValLys 166
 DB 1118 GCCCTGAACACCAACACGACGCGCGCTCAACACCTCAACCGCGCGGTGGCGACAGAC 1177
 QY 166 sLysProThrProThrProProValAlaGlnGlnProAlaProValAlaProProValTh 186
 DB 1178 CTGACCG---CCACGACCGCGCTC-----AACCCGCAACACGTCGCGGTGACGGG 1228
 QY 186 rGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAl 206
 DB 1229 -CGGCGCAACGTCAACCGGCTGCGACGGGGGTGACCTCGGCGACGACACGATCGGGC 1287
 QY 206 aThrAsnProValAlaValArgPheGlyThrAlaThrValAlaGlySerThrValThrse 226
 DB 1288 CAACCTGGCGCGCTTCGGGACCGGTGCTGAGAGGTGGCAACTCGAGCAGACACACAC 1347
 QY 226 rAsnGlyMetTrpPheSerGlyArgAspGlyAsnLeu-----AsnAl 241
 DB 1348 GGGCGCTATCGCGCTGACCGGAGGTACGCGCTGACCGCTCAACACGCGCGGACACG 1407
 QY 241 aSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
 DB 1408 CGTAAACAC---ACGTTGACGCAAGCGACGTAACCGGTAACTCAACACACAC 1464
 QY 259 ---IleValIleGlnHisThr-----AsnGlyPheVa 268
 DB 1465 GGGCGTGAACGTCACCAACACCGCGCGGCGACCGCGCGGTACGCTGCGCGGTTCGCT 1524
 QY 268 lSerSerTyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrG 288
 DB 1525 CAACGCGCTGTGACGATCAACCGACTCT-----GCCGCGCGCTCGGCGACGACGCG 1575
 QY 288 YGlnArgIleAlaSerMetLysAsnGlnProSerGlyValAlaLeuPheGlu 305
 DB 1576 CGGCAAGATCGCACGATCACTGCGGAGCTTGGCGCGCGCACGATCGAC 1627

RESULT 20

US-08-614-377A-6
 ; Sequence 6, Application US/08614377A

Patent No. 5976864
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
APPLICANT: No. 5976864ellini, John F.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF
TITLE OF INVENTION: HETEROLOGOUS
TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSER: Fish & Richardson PC
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,377A
FILING DATE: 12-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/194,290
FILING DATE: 09-FEB-1994
CLASSIFICATION: 435
APPLICATION NUMBER: US 07/895,367
FILING DATE: 09-JUNE-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tsao, Y. Rocky
REGISTRATION NUMBER: 34053
REFERENCE/DOCKET NUMBER: 08106/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Caulobacter crescentus
STRAIN: CB 15
US-08-614-377A-6
Alignment Scores:
Pred. No.: 0.297 Length: 3300
Score: 115.00 Matches: 75
Percent Similarity: 36.48% Conservative: 41
Best Local Similarity: 23.58% Mismatches: 134
Query Match: 6.89% Indels: 69
DB: 2 Gaps: 12
US-10-018-706-2 (1-322) x US-08-614-377A-6 (1-3300)
QY 23 GlyValIleThrTrpCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsnSerThr 42
DB 776 GGGCGCCGTGTGACGACGACGCGGCTGGCTGAACCTGTTCCACCGCCTATCCGTGTCG 835
QY 43 SerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGlnVal 62
DB 836 GGGCGGTGGGGTTCG-----ACCTCTGCTGACCAACGCGGACGACACC 880
QY 63 IleThrAspSerGlnGly-----ValProAsnArgTyrGln 74
DB 881 CTGACGGGACCGCCACACACACACGTTCTGTTGGGGTGAAGTGCAGCGGCGGCGAC 940

QY 75 ValLysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpArgLys 94
DB 941 CTGACCGTGGCGACACCGTAGGCGGCTGTGACCGACCGACGCTGTAAGTGGTGCAC 1000
QY 95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThr 114
DB 1001 GCTGCTCGGTGTACGGCTGTGCGACCGCGGCTGACGATC----- 1039
QY 115 LeuTrpSerGlyAspLeuLysValArgGlyLysSerIleSerSerGlyValAsnThrAla 134
DB 1040 -----TCGGCATGTGAACATG 1057
QY 135 HisThrProSerProValAlaValGlnSer-SerArg-----Pr 147
DB 1058 AACGTGACGTGGCGGCTGGGATCACCTGTGAACGTTCTGGCGGTGACGGGTGACC 1117
QY 147 OProValGlnGlnHisProAlaValGln--LysProThrProProValValVally 166
DB 1118 GCCCTGAACACCAACACACGAGCGGCGGCTCAACCGCTCACCGCGCGCGGACAGAC 1177
QY 166 LysProThrProThrProProValValGlnGlnProAlaProValAlaProValTh 186
DB 1178 CTGACCG---CCAGACCGCGCTC-----AACCGCGAACAACGTCGCGGTGACGGC 1228
QY 186 GLValAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAl 206
DB 1229 -CGGCCAAGTCACCGCTGCGCTGACGGGGGTGACTCTGGGACGACGACGTCGGCGC 1287
QY 206 AtThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrse 226
DB 1288 CAACTCGGCGGCTTCGGGACCGGTGCGTGAAGCGTGCAGAACGACGACGACACAC 1347
QY 226 rAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuLe-----AsnAl 241
DB 1348 GGGCGCTATCGCGCTGACCGGTGTGACCGCGGTGACCGGTGCTGAACGCGCGGACGCG 1407
QY 241 aSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyValaser----- 258
DB 1408 CGTGAACACG---ACGTGAACGACGACGACGACGACGACGACGACGACGACGAC 1464
QY 259 ---IleValIleGlnHisThr-----AsnGlyPheVa 268
DB 1465 GGGCGTACGCTGACCCCAACCGCGCGGACCGCGCGGCTGACGCTGCGGCTGCGGT 1524
QY 268 LserSerTyrIleHisIleLysAspAlaGlnValIleThrGlyAspThrValArgThrG 288
DB 1525 CAACGGCGGCTGTGACGATCACCGACTCT-----GCCGCGGCTGCGGACGACCGC 1575
QY 288 yGlnArgIleAlaSerMetLysAsnGlnProSerGlyValAlaLeuPheGlu 305
DB 1576 CGGCAAGATCCCGACCGCTGACCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1627
RESULT 21
US-09-142-648B-6
Sequence 6, Application US/09142648B
Patent No. 6210948
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
APPLICANT: No. 6210948ellini, John F.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
FILE REFERENCE: 08106/002002
CURRENT APPLICATION NUMBER: US/09/142,648B
CURRENT FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: PCT/CA97/00167
PRIOR FILING DATE: 1997-03-10
PRIOR APPLICATION NUMBER: US 07/614,377
PRIOR FILING DATE: 1996-03-12
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 3300
TYPE: DNA

```

/ ORGANISM: Caulobacter crescentus
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (101)...(3178)
US-09-142-648B-6

Alignment Scores:
Pred. No.: 0.297 Length: 3300
Score: 115.00 Matches: 75
Percent Similarity: 36.48% Conservative: 41
Best Local Similarity: 23.58% Mismatches: 134
Query Match: 6.89% Indels: 69
DB: Gaps: 12

US-10-018-706-2 (1-322) x US-09-142-648B-6 (1-3300)

QY 23 GYVAILLEThThrCysIleLeuAlaGlyCysAlaSerLySProThrTyPAsnSerThr 42
DB 776 GGGCGCCCTGTCGACCGAACAACGGCGCTGGCGTGAACCTTTCACCGCCCTAACCTGCTCG 835
QY 43 SerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnVal 62
DB 836 GGGCGTCTGGCTTCG-----ACCTCTGCTGACCAACCGGCAACGAC 880
QY 63 ILeThrAspSerGlnGly-----ValProAsnArgTyrgln 74
DB 881 CTGACCGGACCGCCCAACAACGACGCTTCGCGGTGAAGTCCGCGCGCTGCGAC 940
QY 75 ValLySglnGlyAspThrValSerLyIleAlaArgTyrglyLeuAsnTrpArgGln 94
DB 941 CTGACCGTGGGACACCTGAGCGGGGTGTCGACCGACGCTCTGAACGTGGGTGCA 1000
QY 95 ILeGlyHisIleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeuThr 114
DB 1001 GCTGTCGCGTTACGGCTCTGCGACCGCGGTGACATC----- 1039
QY 115 LeuTrpSerGlyAspLeuLySValArgGlnArgSerIleSerSerGlyValAsnThrAla 134
DB 1040 -----TCGGGATCGAACAAGAG 1057
QY 135 HisTrpProSerProValAlaValGlnSer-SerArg-----Pr 147
DB 1058 AACGAGACCTCGGCGCTCGATCACCTCGAACAACGCTTCGCGCGCTGACGGTCTGACC 1117
QY 147 GProValGlnGlnHisProAlaValGln--LysProThrProProValAlaValVal 166
DB 1118 GCGCTGAACACCAACACGCGCGCGGTCAAAACGTCACCGCGCGGCGTGGCAGAAC 1177
QY 166 SLYSProThrProThrProProValAlaGlnProAlaProValAlaProProValTh 186
DB 1178 CTGACCG-----CAACGACCGCGCTC-----AAGCGGCAACAAGCTCCCTTCACAGGG 1228
QY 186 rGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrglyProValG 206
DB 1229 -CGCGGCAACGTACCGCTCGCTCAACGGCGTGAACCTGGGCAACGACGCGCGCC 1287
QY 206 AThrAsnProValAlaArgArgPheGlyThrAlaThrValAlaGlySerThrValTh 226
DB 1288 CAACCTGGCGCGCTTCGGGACCGGTGCTGAGCGTGGCAACTCGAGACCAACCAAC 1347
QY 226 rAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIle-----AsnAl 241
DB 1348 GGGCGCTTACCGCGTGAACCGGTGTAACGCGCTGACCGGTGCTCAAAACGCGCGGACAC 1407
QY 241 ASerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
DB 1408 CGTGAACACCC--ACCTTGACCAACGCGACGTGAACCGGTGAACCTTCAGACACAC 1464
QY 259 ---IleValIleGlnHisThr-----AsnGlyPheVal 268
DB 1465 GCGCGGTGACGTACCAACCGCGCGCGCGCGCGCGCGGTGACGCTGCGCGGTGCGGT 1524
QY 268 IserSerTyrlleHisIleLeuAspAlaGlnValLySThrGlyAspThrValArgThrcl 288

```

```

DB 1525 CAACGGCGCTGTGACGATCACCGACTCT-----GCCGCGGCTGGCGCACGACCGC 1575
QY 288 yGlnArgIleAlaSerMetLySAsnGlnProSerGlyAlaAlaLeuPheGln 305
DB 1576 CCGCAAGATCGCCACGCGTCAACCTTGGCGAGCTTGGCGCGCCGCCACGATCGAC 1627

RESULT 22
US-09-311-626B-15
/ Sequence 15, Application US/09311626B
/ Patent No. 6399347
/ GENERAL INFORMATION:
/ APPLICANT: Jorgensen, Per Lina
/ APPLICANT: Schmoor, Kirk
/ APPLICANT: Andersen, Lene No. 6399347boe
/ APPLICANT: Schulten, Martin
/ APPLICANT: Outtrup, Helle
/ TITLE OF INVENTION: No. 6399347el Rhamnogalacturonan Hydrolases
/ FILE REFERENCE: 5572,204-US
/ CURRENT FILING DATE: 1999-05-13
/ PRIOR FILING DATE: 1998-05-08/98
/ PRIOR APPLICATION NUMBER: 60/084,358
/ PRIOR FILING DATE: 1998-05-05
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 15
/ LENGTH: 1896
/ TYPE: DNA
/ ORGANISM: Streptomyces coelicolor
US-09-311-626B-15

Alignment Scores:
Pred. No.: 0.176 Length: 1896
Score: 113.50 Matches: 68
Percent Similarity: 35.04% Conservative: 21
Best Local Similarity: 26.77% Mismatches: 99
Query Match: 6.80% Indels: 66
DB: Gaps: 15

US-10-018-706-2 (1-322) x US-09-311-626B-15 (1-1896)

QY 38 ThrTyAsnSerThrSer-----GlySerGlySerHisArgThrSer 51
DB 1075 ACCAATCTCTCACCACCAACAGCGGCAAGGCTACGACGCGGCAACACACGACTCTCC 1134
QY 52 -----GlySerGlyLeuAlaIleGlySerGlnValIleThr 64
DB 1135 GTCCGCGACGTGACGCGTGAACGCGCGGACGATCTGACGCGGATGCGGTGAC 1194
QY 65 AspSerGlnGlyValProAsnArgTyrglnVal-----LysGlnGlyAsp 79
DB 1195 GAC-----AACGGTACCGCTGTGAGACCAACAGAAAGCGGCGGCGAC 1239
QY 80 ThrVal-----SerLySValAlaGlnArgTyrglyLeuAsnTrpArgGlnIleGly 96
DB 1240 GCCATGACGTGGGACCTGACCCGCTCCGCGCGGCTGAGAGAGTTCAAGTTCAC 1239
QY 97 HisIleAsnAsnLeuAsnSerSerTy-----ThrIleTyThrGlyGln-----Tyr 112
DB 1300 GAGGACGCGTGAAGCCCTGCTGTAACCTGCGGACCGCGGACGCGGCGCAATCTCTCG 1359
QY 113 LeuThr-----LeuTrpSerGlyAsp 119
DB 1360 TCCACCGCGCGGACCGCGGACMAACGCGCGGTCTCCGGGACATCTGTCGGGACG 1419
QY 120 LeuLySValArgGln--ArgSerIleSerSerGlyValAsn--ThrAlaHisTrpPro 138
DB 1420 GCGGCGCGGAGTCTGCTGCTCGCGGAGAGCGGATCCGCAACCCCAAGGACCGT 1479
QY 138 rProValAlaValGln-----SerSerArgProProValGlnGlnHisProAl 154

```

Db 1480 GTCCGACGCCGCAAGCCCTCCAGCGCAACTTCTTCCTGTGGAGCGGACACCGCTC 1539
 Oy 154 aValGlnlysrProthrpProValValVallylsrProthrpProthrpProProva 174
 Db 1540 GGTAACTCTCGACGCGACCCAGCTCGACAAAGTACGACCTCGGGGACACCGCGCTG 1559
 Oy 174 lValGlnlysrProthrpProValValProValThrlGlnlysrProthrpProthrpPro 189
 Db 1600 C-----TCACCGGCTCGGGGCGGCTCCCAACAAG---GCACCAAGCGCACCCCGGCTC 1650
 Oy 190 -----PheAlaThrGly-----SerSerGlyValMetGlnPheAr 201
 Db 1651 CTGCGCGCGCGACATCTCTCGGCGACCTGGCGGAGAGGTCTGTGCGCGACCGTCACACAC 1710
 Oy 201 gTyProValGlyAlaThrAnProValValAlaArgPheGlyThrAlaThrValAlaG 221
 Db 1711 ACAGCCCGCGCATCTACTCCACCCCTTACACACAGG---ACACCCGATACACACCTTC 1767
 Oy 221 ySerThrValThrSerAnGlyMetTrpPheSerGlyArg 234
 Db 1768 CTCACGACGACCCAGTACCGACGCGACCTGGCGTGGACGA 1807

RESULT 23

US-09-308-375-1
 ; Sequence 1, Application US/09308375
 ; Patent No. 6300117
 ; GENERAL INFORMATION:
 ; APPLICANT: Genencor International, Inc.
 ; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
 ; FILE REFERENCE: GC394-PCr
 ; CURRENT FILING DATE: 1999-05-14
 ; EARLIER APPLICATION NUMBER: EP9719636.4
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 7100
 ; TYPE: DNA
 ; ORGANISM: Bacillus subtilis
 US-09-308-375-1

Alignment Scores:
 Pred. No.: 1.36 Length: 7100
 Score: 113.50 Matches: 66
 Percent Similarity: 38.91% Conservative: 55
 Best Local Similarity: 21.22% Mismatches: 142
 Query Match: 6.80% Indels: 49
 Gaps: 11

US-10-018-706-2 (1-322) x US-09-308-375-1 (1-7100)

Oy 22 PheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerIysProThrTyraSer 41
 Db 4303 TACGGTATGTGAACCTCTACCAACTTCTCTGTGTAACCCCTCTCAACGTGTGATCA 4362
 Oy 42 ThrSerGlySerGlySerThrIleArgThrSerGlyGlyLeu----- 56
 Db 4363 TATTCAGCGCAAGTATTCAAGTACATAAATTACAGACGCTAGTAATAATACATGTGACCCT 4422
 Oy 57 AlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyGlnValIys 76
 Db 4423 GCCCTTATGTGACGCTGATATTACAGCAAGATCAGGGTT-----AATGCTAAAGCAGCA 4476
 Oy 77 GlnGlyAspThrValSerIleIleAlaGlnArgTyGlyLeuAsnTrpArgGlnIleGly 96
 Db 4477 TCTGTGTAGGTGCGCATGTGATTAATGCACTGATGCCAGACAGCAAAAGCTTGA 4536
 Oy 97 HisIleAsnLeuAsnSerSerTy-----ThrIleTyThrGly 110
 Db 4537 ---GTAAATTAAGCTTACGATCCCTTATCAAAATGTATGGGTGGAACAAAGTACCTGCC 4593
 Oy 111 GlnTrpLeuThrLeuTrpSerGlyAspLeuIysValArgIuArgSerIleSerSerGly 130

Db 4594 CAACAACCTGAAAGATTGGCGGTAAATGTGAAAAAGCAATGCTGCATATATATATGCGGG 4653
 Oy 131 ValAsnThrAlaIleThrProSerProValAlaValGlnSerSerArgProPro-ValG 150
 Db 4654 -----CCTGTAACTTAATTAATATATATATATATATATATATATATATATATAT 4695
 Oy 150 nGlnIleProValValGlnlysrProthrpProProValValVallylsrProthrpPr 170
 Db 4696 GAAACACAGAAATTACGTCAGAAAGATCATGCGCAACTATATGCAAAATGCTCATTCG-- 4753
 Oy 170 oThrProProValValGlnlysrProthrpProValAlaProProValThrlGlnlysr 190
 Db 4754 -----CAACTTCTTCAATGCGCAGCT-----ATTATACAAATATATAGC 4791
 Oy 190 eAlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAnProva 210
 Db 4792 GCTTTTAGGGTAACTCCAAATATGACAAACAGGAATCTGTCTCCGCTCC--TCCCA-- 4848
 Oy 210 lValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAnGlyMetTr 230
 Db 4849 -----CACAAAGGACTCA 4862
 Oy 230 pPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAl 250
 Db 4863 TTTTGTCGCAAAAGCAGGTACAGCAATTAATCTCTTCAAAAGTGTAAAGTCAAAATTC 4922
 Oy 250 aAspHisAsnMetAsp---GlyAlaSerIleValIleGlnHisThrAnGlyPheValSe 269
 Db 4923 TGGCTACAGTAAACCTCAGATTAACCTGGTGTATTAACACAGATATGGAACAGTTGC 4982
 Oy 269 rSerTyTrIleHis---IleIysAspAlaGlnValIysThrGlyAspThrValArgThrG 288
 Db 4983 CAATTAATGACACATGCTTACACTCTCTGTAAAGCAGGTCAATCATGTTAAAGCCGG 5042
 Oy 288 yGlnArgIleAlaSerMetIysAsnGlnProSerGlyAlaAla---LeuPheGlnPheAr 307
 Db 5043 TCAAACTATTGTAAAGTGTGTAGTACAGGGAATCGACGGAACACCACTTATTACA 5102
 Oy 307 gIleSerArgAnGlyValIlyrValAlaPheTr 317
 Db 5103 GATCGAACAAATGTGAAAAACAAATCATCTCT 5133

RESULT 24

US-08-878-546-9/C
 ; Sequence 9, Application US/08878546
 ; Patent No. 5952463
 ; GENERAL INFORMATION:
 ; APPLICANT: SHIBANO, YUJI
 ; APPLICANT: KIKUCHI, NORIHISA
 ; TITLE OF INVENTION: NOVEL PROTEINASE INHIBITOR AND
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: STEINBERG, RASKIN & DAVIDSON P.C.
 ; STREET: 1140 AVENUE OF THE AMERICAS
 ; CITY: NEW YORK
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/878,546
 ; FILING DATE: 19-JUN-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 158677/1996
 ; FILING DATE: 19-JUN-1996

LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")
US-07-945-283-1

Alignment Scores:
Pred. No.: 2.22 Length: 8438
Score: 112.50 Matches: 40
Percent Similarity: 41.53% Conservative: 9
Best Local Similarity: 33.90% Mismatches: 58
Query Match: 6.74% Indels: 12
DB: 1 Gaps: 3

US-10-018-706-2 (1-322) x US-07-945-283-1 (1-8438)

QY 125 ArgSerIleSerSerGlyValAlaThrAlaHisThr-----ProSerProValAla 141
Db 6438 CGGGCTCTCATCAGGGCTCCGAGCAGGTCACCTCTCTCGGGCCCGGCTCCCGGGCC 6379

QY 142 ValGlnSerSerArgProProVal-GlnGlnHisProAlaValGlnLysProThrPro 161
Db 6378 GGCTCCGGGCGCCGCGCGGCGGACCCCGGCGGAGCTCCGAGCGGAGCGGGGGCCCGCC 6319

QY 161 ovalValValValValValValValValValValValValValValValValValVal 180
Db 6318 GGG 6259

QY 180 ovalAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGln 200
Db 6258 TCCGGCTCTCTCGGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6207

QY 200 eArgTyProValGlyAlaThrAenProVal-ValAArgArgPheGlyThrAlaThrVala 220
Db 6206 -----CCGCGAGGGGAGACCCGCTCGTACGCGGCGGCGGCGGCGGCGGCGGCGG 6154

QY 220 laGlySerThrValThrSerAlaGlyMetTrpPheSerGlyValArgGly 236
Db 6153 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6105

RESULT 26
US-08-923-137-2/c
Sequence 2, Application US/08923137
Patent No. 6083716
GENERAL INFORMATION:
APPLICANT: Wilson, James M.
APPLICANT: Farina, Steven F.
APPLICANT: Fisher, Krishna J.
TITLE OF INVENTION: Chimpanzee Adenovirus Vectors
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr., P.O. Box 457
City: Spring House
STATE: Pennsylvania
COUNTRY: United States of America
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923.137
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,700
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNVN.021CPIUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 36519 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-923-137-2

Alignment Scores:
Pred. No.: 23.9 Length: 36519
Score: 112.00 Matches: 49
Percent Similarity: 39.79% Conservative: 27
Best Local Similarity: 25.65% Mismatches: 83
Query Match: 6.71% Indels: 33
DB: 3 Gaps: 6

US-10-018-706-2 (1-322) x US-08-923-137-2 (1-36519)

QY 26 ThrThrCysIleuAlaGlyCysAlaSerTyProThrTyAenSer-----ThrSerGly 44
Db 7308 TCACGATGTACTTCTTGGCTTGGCCCAACCCCGCTTCAAGAACCGGAGCTTCCAGC 7249

QY 45 SerGlySerHisArgThrSerGlyGlyLeuAlaIleGlySerGlnValIleThr 64
Db 7248 TCTGGAGGAGG-----GGGCTGCCAGATTCGATTCGATTCAGACAGT 7204

QY 65 AspSerGlnGlyValProAlaArgTyGlnValLysGlnValAspThrValSerTyIle 84
Db 7203 TCTCAAGTCATGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7145

QY 85 AlaGlnArgTyGlyLeu-AsnTrpArgIleGlyHisIleAsnAlaSerSer 104
Db 7144 GCCAGGCTTACCCCTTCCCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7085

QY 104 rTyThr-----IleTyThrGly-----110
Db 7084 TTCTACATGCTAGGCTTACCGGAGATCAGAGAGGTTCCCGCTCAAGAGTACTGG 7025

QY 111 -----GlnTrpLeuThrLeuTrpSerGlyAspLeuValArgGluAr 125
Db 7024 AAGAGCGGAGAGTCTCTTCAACCGGAGCTGTGAAC-----AA 6980

QY 125 gSerIleSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSer 145
Db 6979 AAGGACAGCTCAAGATGACATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6920

QY 145 rArgProProValGlnGlnHisProAlaValGlnTyProThrProProValValVala 165
Db 6919 CTAGTCACCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6860

QY 165 lLysLysProThrProThrProProValValGlnGlnProAlaProValaProProVa 185
Db 6859 TCGGCGGAGCTGCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6802

QY 185 lThrGlnAlaProPheAlaThrGlySerSer 195
Db 6801 -ACTGCATGCCATTTCGCGCAGATCGTCT 6772

RESULT 27
US-09-071-035-221
Sequence 221, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and polypeptides
NUMBER OF SEQUENCES: 496

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rni

Page 20

```

CORRESPONDENCE ADDRESS:
ADDRESSSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/03/071,035
FILING DATE:

CLASSIFICATION:
Prior Application DATA:
Application NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brooks
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 867 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-221

```

Alignment Scores:	
Pred. No.:	0.0912
Score:	111.00
Percent Similarity:	34.77%
Best Local Similarity:	24.83%
Query Match:	6.65%
DB:	
	4
	Gaps:
	14
Length:	866
Matches:	75
Conservative:	30
Mismatches:	120
Indels:	77
Gaps:	14

US-10-018-706-2 (1-322) X US-09-071-035-221 (1-867)

```

OY      LeuLeuPegGlyValIleThrTrpCysIleLeuAlaGlyCysAlaSerIleProThrTrp 39
Db      20  :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
OY      31  ATATTATTGCGCTTATTCAGTACAAACGCTTTTGGACGGGGAAGCAGTGTTCGCTTAT 90
Db      40  AAmethrHisSerGlySerGlySerHisValGlyThrSerGlySerGlyLeuAlaIleGly 59
OY      91  GCGCGMAAATCAGAACAAATCTTGTCGAAACAAACAGGAGT----- 132
Db      60  SerGlnValIleThrIleProSerGlnGlyValProAsnArgTrp-----GlnValIleGly 77
OY      133  -----GTTTATACAGATGCAACGAAATGATACCAATGACACCAATAGACCAAGTACAGCA 186
Db      77  nGlyAspThrValSerIleValIleGlnArgTrpGlyLeuAsnTrpArgGlnIleGln 97
OY      187  GAGCAACCAACAGACGACCAAGTACACAGAGC-----AACCATGCAACCGTCAACA 237
Db      97  sIleAsnAspLeu-----AsnSer 104
OY      238  CCAACGCAACTGTAGTAGCGCTTCGAAACCGACGATCCTTGTTCCAGACGAAACCGAGC 297
Db      104  rTyrrThrIleTrpThrGlyGlnTrpLeuThrIleTrpSerGlyAspLeuValAlaArg 124
OY      298  GTRCAAA-----CAGACCAACA 315
Db      124  uArgSerIleSerSerGlyValAlaThrTrpAlaHisThrProSerProValAlaValGln 143
OY      316  AGCCCAAGTAAAGCCAGACGACCAACAGACGACCAAGT-----GTACCGAGACCA 372
Db      144  SerSerArgProProValGlnGlnHisProAla--ValGlnIleProThrProProVal 162

```

```

Db      373  CCAACGAGCCAAAGTGTACCAAGAAAAACCAAGTAACCAAAATTAACCAACCGAGCCA--- 429
      ::: ||||| ::: |||||
Oy      163  ValValValValValValValValProThrProThrProProValValGlnGlnProAlaProValAla_162
      ::: :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      430  -----GAAAGCGCTGTGCCAGTTGTCT-----GAAAAACCGATTGTA 468
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      183  ProProValThrGlnAlaPro-----PheAlaThr 192
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      469  CCAGAACCAACGAGAGCAACCAACAGATGTGTGTGTAAAGCCAAATGAGAAATTCGACCA 528
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      193  GlySerSerGlyValMetGlnPheArgTrpProValGlyValAlaThrSer---ProValAla 211
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      529  GGAGAGATCT---ACACAACGACCAACTGTCTCCATTTGAACCAAGATTAACCTTCACAGAGTA 585
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      212  ArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPhe 231
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      586  ACACATGCCCAACTGTGACGACACCGATTGAACGCAAGC----- 627
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      232  SerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAsp 251
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      628  -----GGGAAAGCAATGTGCGAGTGAGTAAGGGC-----GTCTCTTAACA 669
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      252  HisAsnMetAspGlyValAserIleValIleGlnHisThrAsnGlyPheValSerSerTrp 271
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      670  CAAACGCGTGATGGATTAAACCGATTAAAGTGAATATTAAGATTACCAAGTGGCAAT 729
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Oy      272  IleHisIleIleValAspAlaGlnValIleThrGlyAspThrValArgThrGlyGlnArgIle 291
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      730  GTACACAGTAAAGAGTGTGACGCGAAAAATGAAGTACTTCCTTACCACTGGTGTAAAAAATG 789
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID. NOS: 5674
 SEQ ID NO 1165
 LENGTH: 993
 TYPE: DNA
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-1165

Pred. No.:	0.112	Length:	99
Pragmatic Scores:		Matches:	36
Score:	11.00	Conservative:	17
Percent Similarity:	45.69	Mismatches:	11
Best Local Similarity:	31.03	Indels:	22
Query Match:	6.65	Gaps:	4
DB:	4		

US-10-018-706-2 (1-322) x US-09-134-001C-1165 (1-993)

QY 15 IIEIYARGLDENCILEUHEPHEGIVALLIETHTTHCSYGLILEUHALGLYCSALA 34
 Db 160 ATTGCAATTAATTAACACTTAATGATGCACTTCATTTAAATTCOSCTAATCAAGTA 21
 QY 35 SETLYSPROTHTYAAHSEHTHSEGLISYSEHNIARHTHSEGLYSEGLY 54
 Db 220 TTGAAA-----GTATCAGGCTCATTTCAAGCAACGTCAACAAATKCT 26
 QY 55 GLYLEUHALLEGLYSECDVALLEHTHAPSESGINGLYVALPROAHMARGYRGIN 74

Db 265 GGCAAGCTT-----TATACA 279
 Qy 75 VallysgnglyAspThrValSerlysllealaglnArglyLeuasnTrpArglu 94
 Db 280 GTTAAAGCTGAATTCATCTCTATTCGTCGAAATAGGGTACAACTTATCAAAA 339
 Qy 95 lIeeglyHslleasnleuasnSerlyThrIlelyrThrglyGlnTrpleuthr 114
 Db 340 ATCAATGCAACTTAATGGGTAAAT--AACTATCTTATTTCCCGACAAAGTTGAAA 396
 Qy 115 LeutpSerGlyAspLeuLyValArgluArgSerlleSerSergly 130
 Db 397 GTT--TCGTGAAGCGACGAGTTCACATCGTCAAAAGCTAGTGGG 441

RESULT 29

US-08-217-327-7
 ; Sequence 7, Application US/08217327
 ; Patent No. 5474925
 ; GENERAL INFORMATION:
 ; APPLICANT: John, Maliyakal E
 ; APPLICANT: Barton, Kenneth A
 ; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles and Brady
 ; STREET: P.O. Box 2113
 ; CITY: Madison
 ; STATE: WI
 ; COUNTRY: USA
 ; ZIP: 53701-2113
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/217,327
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/812,233
 ; FILING DATE: 19-DEC-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seay, Nicholas J
 ; REGISTRATION NUMBER: 27,386
 ; REFERENCE/DOCKET NUMBER: 1122990831
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 608-251-5000
 ; TELEFAX: 608-251-9166
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2077 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; IMMEDIATE SOURCE:
 ; CLONE: xylanase gene XYNB
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 299..2077
 ; FEATURE: 299..2077
 ; NAME/KEY: misc_feature
 ; LOCATION: 299..632
 ; OTHER INFORMATION: /function= "cellulose binding
 ; OTHER INFORMATION: domain"
 ; US-08-217-327-7

Alignment Scores: 0.394 Length: 2077
 Pred. No.: 110.50 Matches: 74

Percent Similarity: 34.49% Conservative: 45
 Best Local Similarity: 21.45% Mismatches: 125
 Query Match: 6.62% Indels: 102
 DB: 1 Gaps: 16

US-10-018-706-2 (1-322) x US-08-217-327-7 (1-2077)

Qy 1 MetthrValThrIlealeasnSerGlnasnGlnPro-----llelys 16
 Db 287 ATGAGATTAACATGACAAATTTCCGCAAGCATTAACGTCACCGGGTAATTTTAA 346
 Qy 17 Argleugly-----LeuilepheglyValIleThrThyCysileuAlaglyCysAla 34
 Db 347 CGTACAACAGCGATTGTGGTGGTGGCAGTCGACTTAACAGCCCTGGCTTAAAGCATCG 406
 Qy 35 SerlySproThrlyrYasnSerThrSerglySerglySerHlsarGthrSerglySergly 54
 Db 407 GCTCCCTGTACCTTACACATCGATGAGC-----GAATGTCACCC 445
 Qy 55 GlyleuAlaIleGlySerGlnValIleThrAspSerGlnlyValProasnArglyGln 74
 Db 446 GGCTTTACCGCAATATACCCCTCAAAACGATACCGGTCCCGCATCATTAATCGAAC 505
 Qy 75 VallysgnglyAspThrValSerlysllealaglnArglyLeuasnTrpArglu 94
 Db 506 GTGAATTGGCAATCTCAGCAATCGCATGACAGCGCGGTGAGATGCAACTTCGCGC 565
 Qy 95 lIeeglyHslleasnleuasnSerlyThrIlelyrThrglyGlnTrpleuthr 114
 Db 566 ACCAACCCCTTACAAAGCCACCAACATGAGC----- 595
 Qy 115 LeutpSerGlyAspLeuLyValArgluArgSerlleSer----- 128
 Db 596 ---TGAACGGCAGACATC---GCCCGAGACAAATCTCTCTTCCGCTTCAGGGCGCA 649
 Qy 129 ---SerglyValaenThr-AlaHlsThrProSerProValAlaValGlnSerSerArgPr 147
 Db 650 AAAAATGGCAGCAGCGCGAGCGACCAACCGTCAACCGCGCGCTTGAACAGGCAACC 709
 Qy 147 OProValGlnGlnHslProAlaValGlnlySproThrProProValAlaValAl 165
 Db 710 ACCAGCTCTGTGGCTTCAGCTTCAACCCCA---CCACAGTTCATCTTCGATGCC 766
 Qy 166 -----LyslyPro----- 168
 Db 767 AGTGTGCTCCGCGATGCTGTTGCAAGAACACAGCGGTTTCTGCGCTGAGTGGC 826
 Qy 169 -----ThrProThrProProValAlaGlnInProAlaProValAlaPr 183
 Db 827 ACCATGATTAATACCAACACCGGCTTACCGGAGTGGCTTGGCAACACCAACATGCC 886
 Qy 183 OProValThrGlnAlaProheAla-----ThrglySerSergly-----ValMetG 199
 Db 887 CAGGGG-GCAGCGGTAGTGGGCGATAGATGATACCAAGCATGGCGTCGACCCGTGAC 945
 Qy 199 nPheArglyrProValAlaThrAsnProValAlaArgpHeGlyThrAlaThrVa 219
 Db 946 TATCGGCTATGCCAATGATGGAAC-----GCCAATGCAATGCTCATCTGTGAT 996
 Qy 219 lAlaGlySer-----ThrValThrSerasnGlyMetTrpPhe 232
 Db 997 TAAAGGGGCGACGAGTAAGTATACGTGAGATTGCCACACACCGCGCTGAGACAC 1056
 Qy 232 rGlyArg-----AspGlyAspLeuIleasnIleSerAsn-----Al 244
 Db 1057 CTGGCAAAACCGCACTATGATGATGGATTTGGTACAGCGCAATATATTTGCGAGTTGTC 1116
 Qy 244 aglyThrValIleGlnAlaAspHlsAsnMetAspGlyAlaSerlleValIleGlnHsl 264
 Db 1117 TGCACGACAGCGCAAGGCTTACCCCAATATGATTCGTTAAGTGTGTG----- 1165
 Qy 264 rAsnGlyPheValSerSerglyrIleHsllelyAspAlaGlnVallyThrGlyAspThr 284

Db 1166 -----GGTGGTAC 1173

QY 284 TValaThrgly 288

Db 1174 GGTACGGCGGGT 1186

RESULT 30

US-09-136-574A-2

Sequence 2, Application US/09136574A

Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESSES:

ADDRESS: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 6416 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-136-574A-2

Alignment Scores:

Prod. No.: 2.26 Length: 6416

Score: 110.50 Matches: 74

Percent Similarity: 33.96% Conservative: 34

Best Local Similarity: 23.27% Mismatches: 99

Query Match: 6.62% Indels: 111

Db: 4 Gaps: 18

US-10-018-706-2 (1-322) x US-09-136-574A-2 (1-6416)

QY 51 SerGlySerGlyGlyLeuAlaIleGlySer-----Gln 61

Db 4003 AGTGAGCGGTGTTTGGACAGCGGACAGACAGAGATAGGCTTAGCATACAG 4062

QY 62 ValIleThrAspSerGlnGlyValProAsnArgTYrGlnValIleGlnGlyAspThrVal 81

Db 4063 AAGGCACTGGCAGCTACATCAATGACTATTCGTAGAAAGT----- 4110

QY 82 SerIlyIleAlaGlnArgTYrGlyLeuAsnThrArgIleGlyIleAsnAsnLeu 101

Db 4111 -----GGACAGCGCTATATAGAGAAACAGAAAGTAAACAGGCTATATAGATGT--- 4158

QY 102 AsnSerSerTYrThrIleTYrThrGlyGlnThrLeuThrLeuTPSerGlyAspLeuIys 121

Db 4159 -----CGATAGTGTGGGA----- 4173

QY 122 ValArgIleArgSerIleSerSerGlyValAsnThrIleAsn-----ThrProSerPro 139

Db 4174 -----AGAGAGCCGAGCAGGGGTACAAAGCCGCGGAGGAGAGTGAACCCAGACACCG 4224

QY 140 ValAlaValGlnSerSerArgProProValGlnIleHisProAlaValGlnIleProThr 159

Db 4225 GCACCGAGCCGACATCGACCA-----ACACCAACA 4257

QY 160 ProProValValValValValValValProThrProThrProProValValGlnIleProAla 179

Db 4258 CTACACACCA-----CCGACACGACCGACGCTGACGTCGACCCCACT 4305

QY 180 ProValAlaProProValThrGln----- 187

Db 4306 CCT--ACACCTGGGGTAAACCCGATGTAAATATCATCATCATCGTCGACGGGAGAA 4362

QY 188 -----AlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTYrProVal 204

Db 4363 ACAGAGATTAACCCGCTATATTATAGAGCAATCGAGATATCCAG----- 4407

QY 205 GlyAlaThrAsnProValValValArgArgPheGlyThrAlaThrValAlaGlySerThrVal 224

Db 4408 GGTGTGTTCACACT--GCMAAGCAGCTTGT-----GGAAACAGATYG 4449

QY 225 ThrSerAsnGlyMetTrpPheSerGlyArgAspIlyAspLeuIleAsnAlaSerAspAla 244

Db 4450 ACGGTTTACAT--TGGAGAAC-----ANTATGTCATGCA 4485

QY 245 GlyThr-----ValIleGln 249

Db 4486 GGAAGTACGTGTATCATTCACAGCATATATATGTTATATATATGATGATATACAGGG 4555

QY 250 AlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis-----Thr 264

Db 4546 AATGATTAAGAACCTTCAGACGCTGTGTAGCAAAATTCACAGCAGCATTAACAA 4605

QY 265 AsnGlyPheValSerSerTYrIleIleIleIleAspAlaGlnValIlyThrGlyAsp--- 283

Db 4606 AATGCATATTCAGGCATCAATTCAGATGTAGTGTATGTGCGAAAGATGGAAATGT 4665

QY 284 ThrValArgThrGlyGln-----ArgIleAlaSerMetIysAsnGlnPro 298

Db 4666 ACAGTGAAGCAGTCAAGACAGCTCCGTCGCCAGAGAGGCTAGGCTTAAGTTAAAAA 4725

QY 299 SerGlyAlaIleAsnLeuPheGlnPheArgIleSerArgAsnGlyValTYrValAsp 316

Db 4726 GATGTGCACTGTATTCAGAGCTGACGTGAATGATTAATATATATATGAT 4779

RESULT 31

US-09-136-574A-46

Sequence 46, Application US/09136574A

Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-AUG-1998
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 2029 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-136-574A-46

Alignment Scores:
Pred. No.: 0.475 Length: 2029
Score: 109.50 Matches: 58
Percent Similarity: 33.47% Conservative: 24
Best Local Similarity: 23.67% Mismatches: 80
Query Match: 6.56% Indels: 83
DB: 4 Gaps: 13

US-10-018-706-2 (1-322) x US-09-136-574A-46 (1-2029)
Qy 34 AlaserlypsrothTyraSn-----SerThrseryGlySerglySerThrsArg 49
Db 193 GCAAGCATGTGACATTTGTTGTAAGCTTAGACGCGAGTGACGCGATTTAT 252
Qy 50 -----ThrseryGlySerglySerglySerglySerglySerglySergly 63
Db 253 TACCTGAGGTAGATTGACAGTGTGAGCTGG----- 285
Qy 64 ThrserseryGlyValProAsnArgTyrgInvallySgInglyAspThrValserly 83
Db 286 -----CAgTTTCAGCGCTGTGAAGACACAGCGGAT 315
Qy 84 lleaaglnArgTyrglyLeuasn-----TryArg 93
Db 316 ATACAGGTAAAGTTTAAACAAATGACGAGCAATTAATCAAGCAGCAGCTGTGCA 375
Qy 94 GluilegylHsileAsnAsnleu-----AsnserSeryTyrrHlleTyrrHclgln 111
Db 376 TGGTTCAGACATGACAAATTATGAGAGAGAGAGGAGCGGTGTATGTA--GAT 432
Qy 112 TryleuThrleuTyrSerglyAspleuTyValArgIuarSerrIleserseryGly 131
Db 433 GGTGTTCGTATGCGG-----CAGAGCCCGGAGCAGCGGTG 471
Qy 132 AsnThralaHsThrProSerProValAlaValGlnSerSerrArgProProValGln 151
Db 472 ACCCAACTTCTACACCCACACCGGTTTCATCATCTCACT----- 510

Qy 152 HisProAlaValGlnlypsrothProthProValValVallySelysProthProth 171
Db 511 -----CTTACACCAACAGCAACG-----CCACACCTTACA 540
Qy 172 ProProValValGlnGlnProAlaProValAlaProProValThrgIuaIaPro----- 189
Db 541 CCTTCTATCAGATTAACACCGCGCCAACTGCAACACCC--ACTCCGACTCTTCTGTC 597
Qy 190 -----PheAlaThrglySerseryGlyValMetGlnPhe 200
Db 598 ACAGATATCAAAATGATGATTGTTATTTGCGAGGTAAACAAATATGACACAGAT 657
Qy 201 ArgTyProValGlyAlaThrsAsnProValValAlaArgSphgelyThralaThrValAla 220
Db 658 GGCAACTGTATGTGTTAACAAG-----GTTATGTGTTGATTTAATCA----- 705
Qy 221 GlySerThrValThrsersenglyMetTyrPheSeryIyArgAspGlyAspleuileAsn 240
Db 706 ---GGAACGATGTGTGATGTGTGTGAGTTGTAATCTTAAAGTCATTAAGCTGAG 762
Qy 241 AlaserAsnAlagly 245
Db 763 ATTGCAACAGAGCA 777

RESULT 32
US-08-917-320-18
Sequence 18, Application US/08917320
Patent No. 5824508
GENERAL INFORMATION:
APPLICANT: Spaele, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: No. 5824508 Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/917,320
FILING DATE: 25-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Luann Gser
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
US-08-917-320-18
Alignment Scores:

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rn1

Page 24

Pred. No.: 1.42 Length: 3833
Score: 109.00 Matches: 66
Percent Similarity: 33.68% Conservative: 30
Best Local Similarity: 23.16% Mismatches: 116
Query Match: 6.53% Indels: 74
DB: 1 Gaps: 13

US-10-018-706-2 (1-322) x US-06-917-320-18 (1-3833)

```
QY 35 SerIyPProThrThyAsnSerThrserySerIySerHsArgThrSeryIySeryIy 54
DB 2301 ACCTCCCTTACCTTGAATACATGATTTGCTGATCCCAATACACACAGGT----- 2354
QY 55 GlyLeuAlaIleGlySerGlnValIleThrsPserGlnGlyValProAsnArgTyr--- 73
DB 2355 -----CTACCCAGCTTACTCACTGCTCCACCACTCACC 2390
QY 74 ---GlnValIyGlnGlyAspThrValSerIyValIleAlaGlnArgTyrGlyLeuAsnTrp 92
DB 2391 GCACCTGCACACAGGCCCCCTGTATCC----- 2420
QY 93 ArgGlnIleGlyHsIleAsnAsnLeuAsnSerSeryTyrThrIleTyrThrGlyGlnTrp 112
DB 2421 -----ACCGCGATGTCAACCCCAACACACAGCGGACAGTCAGCGCATCACC 2474
QY 113 LeuThr-----LeuTrpSeryIyAspLeuValArgGlnArgSeryIleSer 128
DB 2475 GTGACACCAAGTCATCTCATGGACACACGACAGAAAGTAAGGCCCCGACATGACC 2534
QY 129 SerGlyValAsnThrAlaHsThrsPserProValAlaValGlnSerySeryArgPropo 148
DB 2535 AGCTCCACCTCAGCAGTACTACCCCAACCCCAATGTCACC---AGCCCCACCCAGCA 2591
QY 149 Val-----GlnHsIleProAlaValGlnHsProThrPro 160
DB 2592 GTGACTACCCCAACCCCAATGTCACACAGGCCCCCACTGACTACCCCAACCCCA 2651
QY 161 -----ProValValValIlySlyProThr 169
DB 2652 AATGCCACACAGCCCCCTTGGGAAACAAAGTCTTACCTCCAGACAGTACCCCAACCC 2711
QY 170 Pro-----ThrsProValIyGlnGlnProAlaProValAlaProProValIyGln 187
DB 2712 CCAATATGCCACAGGCCCCCTTGGGAAACAAAGGCCCCCACTCAGCAGTACTACCCCA 2771
QY 188 AlaProPheAlaThrsSerySeryValMetGlnPheArgTyrProValIyAlaThr 207
DB 2772 ACCCCAAATGCCACAGGCCCC-----TTGGGAAACAA 2807
QY 208 AsnProValValArgArgPheGlyThrAlaThrValAlaGlySeryThrValIySeryAsn 227
DB 2808 AGCCCC-----ACCTCAGCAGTACTACCCCAACCCCAATGTCACC 2849
QY 228 GlyMetTrpPheSeryIyArgPheGlyAspLeuIleAsnIleSeryAsn-----Ala 244
DB 2850 GGGCCCTACTGTG---GAGAAACACAGTCACAGGCAAAATGCCACACACACCTTAGA 2906
QY 245 GlyThr-----ValIleGlnAlaAspHsAsnMetAspGlyAlaSeryIle 259
DB 2907 GGAACAGTCCACACAGTACTAGTACACAGCAACCAAAATGCAACACAGCTGTATACC 2966
QY 260 ValIleGlnHsIleThrsSeryIyPheValSeryTyrIleHsIleValAspAlaGlnVal 279
DB 2967 ACAGGCAACATTA-CATTAATCTCAAGTTCAACCTTCCATGTCACTGACAGACCAAGTTC 3025
QY 280 LysThrGlyAspThr 284
DB 3026 AAACCCAGAGACACT 3040
```

RESULT 33
PCT-US95-04611A-18
Sequence 18, Application PC/TUS9504611A
GENERAL INFORMATION:

APPLICANT: Spaete, Richard and Jackman, Winthrop T.
TITLE OF INVENTION: Non Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
CITY: Palo Alto
STREET: 5 Palo Alto Square
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patgen, In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04611A
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Luanm Geert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELEPHONE: 415-857-0663
TELEFAX: 415-843-5163
TELEX: 380816 COOLEYPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
PCT-US95-04611A-18

Alignment Scores:
Pred. No.: 1.42 Length: 3833
Score: 109.00 Matches: 66
Percent Similarity: 33.68% Conservative: 30
Best Local Similarity: 23.16% Mismatches: 116
Query Match: 6.53% Indels: 74
DB: 5 Gaps: 13

US-10-018-706-2 (1-322) x PCT-US95-04611A-18 (1-3833)

```
QY 35 SerIyPProThrThyAsnSerThrserySerIySerHsArgThrSeryIySeryIy 54
DB 2301 ACCTCCCTTACCTTGAATACATGATTTGCTGATCCCAATACACACAGGT----- 2354
QY 55 GlyLeuAlaIleGlySerGlnValIleThrsPserGlnGlyValProAsnArgTyr--- 73
DB 2355 -----CTACCCAGCTTACTCACTGCTCCACCACTCACC 2390
QY 74 ---GlnValIyGlnGlyAspThrValSerIyValIleAlaGlnArgTyrGlyLeuAsnTrp 92
DB 2391 GCACCTGCACACAGGCCCCCTGTATCC----- 2420
QY 93 ArgGlnIleGlyHsIleAsnAsnLeuAsnSerSeryTyrThrIleTyrThrGlyGlnTrp 112
DB 2421 -----ACCGCGATGTCAACCCCAACACACAGCGGACAGTCAGCGCATCACC 2474
QY 113 LeuThr-----LeuTrpSeryIyAspLeuValArgGlnArgSeryIleSer 128
DB 2475 GTGACACCAAGTCATCTCATGGACACACGACAGAAAGTAAGGCCCCGACATGACC 2534
```

```

Qy 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
Db 2535 AGCTCACTCACTCACTGAGTACTACCCCAACCCCAATGACACC--AGCCCCACCCCAACA 2591
Qy 149 Val-----GlnGlnHisProAlaValGlnLysProThrPro 160
Db 2592 GTGACTACCCCAACCCCAATGCCACGAGCCCAACCCCAAGTACTACCCCAACCCCA 2651
Qy 161 -----ProValValValValValValValValValValValValValValValVal 169
Db 2652 AATGCCACGAGCCCACTTGAGAAAAACAAGTCTTACTCTACGAGTACTACCCCAACC 2711
Qy 170 Pro-----ThrProProValValGlnGlnProAlaProValAlaProProValThrGlu 187
Db 2712 CCAATATGCCACGAGCCCACTTGAGAAAAACAAGTCTTACTCTACGAGTACTACCCCA 2771
Qy 188 AlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThr 207
Db 2772 ACCCCAAATGCCACGAGCCCACT-----TTGGGAAAAACA 2807
Qy 208 AsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsn 227
Db 2808 AGCCCC-----ACCTCAGCAGTACTACCCCAACCCCAATGCCAC 2849
Qy 228 GlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn-----Ala 244
Db 2850 GGCCCTACTGTG--GGAGAAACAAGTCCACAGGCAATGCCACCAACACACCTTAGA 2906
Qy 245 GlyThr-----ValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259
Db 2907 GGAACAAATGCCACGAGTACTACGCAACCAAAAAATGCAACCAAGTCTTACC 2966
Qy 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleAspAlaGlnAla 279
Db 2967 ACAGGCCAACTTAA-CATTAATTCAAGTTCAACTCTTCCAGTACTAGAGCCCAAGTTC 3025
Qy 280 LysThrGlyAspThr 284
Db 3026 AAACCCAGAGACACT 3040

```

RESULT 34

```

US-08-783-774-1
Sequence 1, Application US/08783774
Patent No. 6054130
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NON-SPLICING VARIANTS OF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edwards
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTA/FASTA Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/783,774
FILING DATE: 15-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7582-037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864

```

```

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1014...3734
OTHER INFORMATION:
US-08-783-774-1

```

Alignment Scores:

```

Pred. No.: 2.79 Length: 5931
Score: 109.00 Matches: 66
Percent Similarity: 33.68% Conservative: 30
Best Local Similarity: 23.16% Mismatches: 116
Query Match: 6.53% Indels: 74
DB: 3 Gaps: 13

```

US-10-018-706-2 (1-322) x US-08-783-774-1 (1-5931)

```

Qy 35 SerLysProThrTyrAsnSerThrSerGlySerHisArgThrSerGlySerGly 54
Db 2301 ACCCTCCCTACTCTGATTAACAGTCTTGCTGATCCCAATACAGCAGAGT----- 2354
Qy 55 GlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyr-- 73
Db 2355 -----CTACCAAGCTCTACTCAGTGGCTTACCAACCTCACC 2390
Qy 74 ---GlnValysGlnGlyAspThrValSerIleLeuAlaGlnArgTyrGlyLeuAsnTrp 92
Db 2391 GCACCTGCAGACAGAGGCCCTCAGTATCC----- 2420
Qy 93 ArgGluIleGlyHisIleAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112
Db 2421 -----ACCGCGATGTCCACGAGCCCAACACGAGCCGACACGTCAGCGCTCACCG 2474
Qy 113 LeuThr-----LeuTrpSerGlyAspLeuValArgGluArgSerIleSer 128
Db 2475 GTGACACCAAGTCCATCTCCATGGGACAGGACAGAAAGTAAGGCCCGCATGAGC 2534
Qy 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
Db 2535 AGCTCACTCACTCACTGAGTACTACCCCAACCCCAATGCCAC--AGCCCCACCCCAACA 2591
Qy 149 Val-----GlnGlnHisProAlaValGlnLysProThrPro 160
Db 2592 GTGACTACCCCAACCCCAATGCCACGAGCCCAACCCCAAGTACTACCCCAACCCCA 2651
Qy 161 -----ProValValValValValValValValValValValValValValValVal 169
Db 2652 AATGCCACGAGCCCACTTGAGAAAAACAAGTCTTACTCTACGAGTACTACCCCAACC 2711
Qy 170 Pro-----ThrProProValValGlnGlnProAlaProValAlaProProValThrGlu 187
Db 2712 CCAATATGCCACGAGCCCACTTGAGAAAAACAAGTCTTACTCTACGAGTACTACCCCA 2771
Qy 188 AlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThr 207
Db 2772 ACCCCAAATGCCACGAGCCCACT-----TTGGGAAAAACA 2807
Qy 208 AsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsn 227
Db 2808 AGCCCC-----ACCTCAGCAGTACTACCCCAACCCCAATGCCAC 2849
Qy 228 GlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn-----Ala 244
Db 2850 GGCCCTACTGTG--GGAGAAACAAGTCCACAGGCAATGCCACCAACACACCTTAGA 2906
Qy 245 GlyThr-----ValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259

```

Db	2307	GGAACAAGTCCACCCACGAGTATTCACGACCAACCAAAAAATGCAACCAAGTGTGTAC	2966
QY	260	ValIeGlnIsthrAsnGlyPheValSerSerTyrIleHISileIyAspAlaGlnVal	279
Db	2367	ACAGGCCCAACATAA-CATTACTTCAAGTTCAACCTTTCATGTCACGAGACCCAGTTC	3025
QY	280	LySThrGlyAspThr	284
Db	3026	AAACCCAGAGACACT	3040

RESULT 35
 US-09-556-706B-1
 Sequence 1, Application US/09556706B
 Patent No. 6458364
 GENERAL INFORMATION:
 APPLICANT: Spaete, Richard
 APPLICANT: Jackman, Wlithrop
 TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
 FILE REFERENCE: 7682-050-9999
 CURRENT APPLICATION NUMBER: US/09/556, 706B
 CURRENT FILING DATE: 2000-04-24
 PRIOR APPLICATION NUMBER: 08/783, 774
 PRIOR FILING DATE: 1997-01-15
 PRIOR APPLICATION NUMBER: 08/229, 291
 PRIOR FILING DATE: 1994-04-18
 NUMBER OF SEQ ID NOS: 19
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 1
 LENGTH: 5931
 TYPE: DNA
 ORGANISM: Virus
 FEATURE:
 OTHER INFORMATION: GP350/220
 US-09-556-706B-1

Alignment Scores:		
Pred. No.:	2.79	Length: 5931
Score:	109.00	Matches: 66
Percent Similarity:	33.68%	Conservative: 30
Best Local Similarity:	23.16%	Mismatches: 116
Query Match:	6.53%	Indels: 74
DB:	4	Gaps: 13

QY	35	SerIysProPrrTyranSerThSerdlySergIysSerHisArgThrSerdlySerdly	54
Dd	2301	ACGtCCCTACTTGAAATACAACAGTGGATTGTGTGAATGCCAATACACACAGCT	2354
QY	55	GlyLeuMaLaIleGlySergInValILeThrAspSergInGlyValProAnaTrgYr---	73
Dd	2355	-----CTACCAGCCTTACTACGTCGCTCACCACTCAC	2390
QY	74	---GlnValIysGlnGIYAAPThValSerIylsIleaIgaIngrglyIleuaSnTrp	92
Dd	2391	GCACTGCACAGACAGGCCCATCTATCC-----	2420
QY	93	ArgGlutIleGlyHISileuAnMaIeuAnsSerTyrrThrlIetyrrThnGIelntRp	112
Dd	2421	-----ACCGGAGATGATACCAAGGCCAACACAGCGGACACACGTACAGCGCATCACCG	2474
QY	113	Leuthr-----LeuTrpSerdlyAspLeuysValArGIuArSeriIeser	128
Dd	2475	GTGACACCAAGTCCATCTCCATGGACACAGCGCACAGMAATAAGCCCCGCAATAGAC	2534
QY	129	SerGIyValenThralahISethProserProValalavalGlnSerSerArXproPro	148
Dd	2535	AGCTCCACTCACAGTGACTTAACCCCAACCCCAATATGCACC--AGCCCCACCCACGA	2591
QY	149	ValI-----GlnInHISProAlaValGlnIysProThPro	160
Dd	2592	GTAGCTACCCCAACCCCAATATGCACACAGCCCCACCCGACAGAGACTTAACCCCAACCCA	2651

[illegible]

RESULT 36
 US-08-314-309A-1
 Sequence 1, Application US/08314309A
 Patent No. 5677141
 GENERAL INFORMATION:
 APPLICANT: ISOGAI TAKAO
 APPLICANT: FURUKAWA, MASAO
 APPLICANT: IKAMI, MORITA
 APPLICANT: ARAMORI, TOSHIRO
 APPLICANT: KOJO, HITOSHI
 TITLE OF INVENTION: PROCESS FOR PRODUCING 7-AMINOCEPHEM
 TITLE OF INVENTION: COMPOUND OR SALTS THEREOF
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, WAIR & NEUSTADT
 ADDRESSEE: P.C.
 STREET: 1755 S. Jefferson Davis Highway, Suite 400
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/314,309A
 FILING DATE: 30-SEP-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/631,906
 FILING DATE: 21-DEC-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Oblon, No. 5677141man F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 18-863-0 CONT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 413-3000
 TELEFAX: (703) 413-2220

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rn1

Page 28

41 -----SerThrSerGlySerGlySerHisArgThrSerGlySer-----G1 54
101 CGGACCGGCTCCCGGAGCCCGGGGGCGGCTCGGACAGATTCAAGCCCATCTTTC 160
54 yGlyLeuAlaIleGlySerGlnValIle-----ThrAspSerGlnGly-ValProAsn 72
161 TGACCCGAGGTTGGGCGGCGGCTCGGCGGTAACAAATCCTAAGAGGTAGCGAGT 220
72 rGlyArgIn-----ValIleGlnGlyAspThrValSerIle-----IleA 85
221 GCTACGAGCGCTTCAGTCCGACAGACAGCGGGGCGGAGATCGTCACTGTG 280
85 laGlnArgTyrGlyLeuAsnTPrArgGlnIleGlyHis-----IleAsn 100
281 TGACCGCGGCGCTGATCCGACCGGAGGCTGGGCGGACAGAAAGTCCAGATCGAGC 340
100 snLeuAsn-SerSerTyrThrIleTyrThrGlyGlnTPrLeuThrLeuTPrSer----- 117
341 AGATGTGAGTGTGTGAGAAACCGACGCGGAGGTGACACGACGTTGAGCTGTTCG 400
118 -----Gly 118
401 AGGCGCAGCAGAGCTGGGCGACACAGTGGGCAACGCGGCAAGGTTGGCGGACAGGC 460
119 AspLeuLysValArgGlnArgSerIleSerSerGlyValAsnThrAlaHisThrProSer 138
461 CCAATGGCGATCGGTAGCGCAGTCTGACCAACGCAAGGAGGCTCAAGCGGCGAGC 520
139 -----ProValAlaValGlnSerSerArgProProVal 149
521 GCAACACAGAAACCGTGAACCGGCTCCAGACACACAGACAGCAGCGCGCT--- 577
150 GlnGlnHisProAlaValGlnLysProThrProProValValValLysProThr 169
578 CGGCGACACCCAGAGAGAAAGAGGCAAGCTTCCA-----AGAAGAAAGAGCGCT 628
170 ProThrPro-----ProValValGlnGlnProAlaProValAlaThrGlnAla 188
629 CCAAGGCGCAGGCGGAGGAGGAGGCTCCCTCCGACCTCCCATCCAGAC----- 679
189 ProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsn 208
680 -----CCAACGAAACCCAGTACTGTCTGTGCAACGAGGTCTCTAGGAGAGATG 733
209 ProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerArgGly 228
734 GCT-----GCGACACAGCAGAGTGGCCCATCGAGTGTTCACCTCTCGT 778
229 MetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle 248
779 GCGTGT-----GGCTCAATCATTAACCAAGGAGCAAGTGTACTGTCCA 823
249 GlnAla 250
824 AGTGCC 829

RESULT 38
; Sequence 9, Application US/08751230
; Patent No. 6117633

GENERAL INFORMATION:

APPLICANT: Garkavtsev, Igor

APPLICANT: Rishovol, Karl

TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Diane, Swecker & Mathis

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/751,230
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mool, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-144
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2061 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURES:
NAME/KEY: CDS
LOCATION: 16..900
US-08-751-230-9
Alignment Scores:
Pred. No.: 0.759 Length: 2061
Score: 107.50 Matches: 66
Percent Similarity: 38.03% Conservative: 42
Best Local Similarity: 23.24% Mismatches: 102
Query Match: 6.44% Indels: 74
Gaps: 14
US-10-018-706-2 (1-322) x US-08-751-230-9 (1-2061)
22 PheGlyValIleThrThrCysIleLeuAlaGly-CysAlaSerLysProThrTyrAsn-- 40
41 TTCCAGATATAGCAGTACGATGCCGCGCTGTGGCTCGGGGCGGCGTGCAGTT 100
41 -----SerThrSerGlySerGlySerHisArgThrSerGlySer-----G1 54
101 CGGACCGGCTCCCGGAGCCCGGGGGCGGCTCGGACAGATTCAAGCCCATCTTTC 160
54 yGlyLeuAlaIleGlySerGlnValIle-----ThrAspSerGlnGly-ValProAsn 72
161 TGACCCGAGGTTGGGCGGCGGCTCGGCGGTAACAAATCCTAAGAGGTAGCGAGT 220
72 rGlyArgIn-----ValIleGlnGlyAspThrValSerIle-----IleA 85
221 GCTACGAGCGCTTCAGTCCGACAGACAGCGGGGCGGAGATCGTCACTGTG 280
85 laGlnArgTyrGlyLeuAsnTPrArgGlnIleGlyHis-----IleAsn 100
281 TGACCGCGGCGCTGATCCGACCGGAGGCTGGGCGGACAGAAAGTCCAGATCGAGC 340
100 snLeuAsn-SerSerTyrThrIleTyrThrGlyGlnTPrLeuThrLeuTPrSer----- 117
341 AGATGTGAGTGTGTGAGAAACCGACGCGGAGGTGACACGACGTTGAGCTGTTCG 400
118 -----Gly 118
401 AGGCGCAGCAGAGCTGGGCGACACAGTGGGCAACGCGGCAAGGTTGGCGGACAGGC 460
119 AspLeuLysValArgGlnArgSerIleSerSerGlyValAsnThrAlaHisThrProSer 138
461 CCAATGGCGATCGGTAGCGCAGTCTGACCAACGCAAGGAGGCTCAAGCGGCGAGC 520

QY 139 -----ProVal1aVal1GlnSerSerArgProProVal 149
Db 521 GCAACACGAGACCGTGAACCGCTCCAGAACACGACACGACGCGCT--- 577
QY 150 GlnGlnHisProAlaVal1GlnLysProThrProProVal1Val1Val1LysLysProThr 169
Db 578 CGGGACACCCACGAGAGAGAGGCAAGACTCCA-----AGAAAGAAAGCGCT 628
QY 170 ProThrPro---ProVal1GlnGlnProAlaProVal1AlaProProVal1ThrGluAla 188
Db 629 CCAAGGCCACGAGCGGAGCGAGCGCTCCCTCCGACCTCCCACTCCACC----- 679
QY 189 ProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProVal1GlyAlaThrAsn 208
Db 680 -----CCAACGACCCACGACGACTCTCTGTGCAACCAAGCTCTCTATGCGAGATGATCG 733
QY 209 ProVal1Val1ArgArgPheGlyThrAlaThrVal1AlaGlySerThrVal1ThrSerAsnGly 228
Db 734 GCT-----GCCGACACGACGAGTCCCATCGAGTGTCCACTTCTCGT 778
QY 229 MetTrpPheSerGlyArgAspGlyAspLeu1LeuAsn1AsnAlaGlyThrVal1Le 248
Db 779 GCGTGG-----GGCTCAATCATTAACCAAGGCGCAAGGACTGTCTCCCA 823
QY 249 GlnAla 250
Db 824 AGTGC 829

RESULT 39
US-09-499-082-9
Sequence 9, Application US/09499082
Patent No. 6143522
GENERAL INFORMATION:
APPLICANT: Helbing, Caren C.
APPLICANT: Klabowol, Karl
APPLICANT: Johnston, Randall N.
APPLICANT: Garkavtsev, Igor
TITLE OF INVENTION: METHODS OF MODULATING APOPTOSIS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09499,082
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/828,158
FILING DATE: 27-MAR-1997
APPLICATION NUMBER: US 08/751230
FILING DATE: 15-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Wool, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-148
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-854-7400
TELEFAX: 650-854-8275
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2061 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 16..900
US-09-499-082-9

Alignment Scores:
Pred. No.: 0.759 Length: 2061
Score: 107.50 Matches: 66
Percent Similarity: 38.03% Conservative: 42
Best Local Similarity: 23.24% Mismatches: 102
Query Match: 6.44% Indels: 74
Gaps: 14

US-10-018-706-2 (1-322) x US-09-499-082-9 (1-2061)

QY 22 PheGlyVal1IleThrThrCysIleLeuAlaGly-CysAlaSerLysProThrTyrAsn-- 40
Db 41 TTCCAGATATGACATGACATGATCCCGGCTGTGCTCGGGCCGGGCTGCAGTT 100
QY 41 -----SerThrSerGlySerGlySerHisArgThrSerGlySer-----G1 54
Db 101 CGACCGCGCTCCGCGACCGCGCGCGCGCTCGAGACATTCAGGCGCATCTTGC 160
QY 54 YGlyLeuAla1IleGlySerGlnVal1le-----ThrAspSerGlnGly-ValProAsn 72
Db 161 TGACCCGAGGGGTGGCGCGCGCGCTGCGAACAAGATCTTGAGAGCTTAGACGAGT 220
QY 72 rGlyrGln-----Val1GlnGlyAspThrValSerLys-----IleA 85
Db 221 GCTACGAGCGCTTCACTGCGAGACAGACGCGCGCGAGAGCGGAGTCTGCACTGTG 280
QY 85 IaglnArgTyrGlyLeuAsnTrpArgGluIleGlyHis-----IleAsn 100
Db 281 TGCAGCGCGCGCTGATCCGACCGAGAGCTCGGCGCGAGAAATCCAGATCGTAGCC 340
QY 100 snLeuAsn-SerSerTyrThrIleTyrThrGlyGlnTrpLeuThrLeuTyrSer----- 117
Db 341 AGATGTGTGAGCTGTGTGAGAACCGCACGCGCGAGGTGACACGACGAGCTGTTCG 400
QY 118 -----Gly 118
Db 401 AGCGCGACGAGAGCTGGCGGACACAGTGGGCAACCGCGCAAGTTGGCGCGACGCG 460
QY 119 AspLeuLysVal1ArgGluArgSerIleSerSerGlyVal1AsnThrAlaHisThrProSer 138
Db 461 CCATGGCGATCGCGTGGCCGATCTGACAAAGCCACACGACGAGCTCACGGCGGACG 520
QY 139 -----ProVal1aVal1GlnSerSerArgProProVal 149
Db 521 GCAACACGAGACCGTGAACCGCTCCAGAACACGACACGACGAGCGCGCT--- 577
QY 150 GlnGlnHisProAlaVal1GlnLysProThrProProVal1Val1Val1LysLysProThr 169
Db 578 CGGGACACCCACGAGAGAGAGGCAAGACTCCA-----AGAAAGAAAGCGCT 628
QY 170 ProThrPro---ProVal1GlnGlnProAlaProVal1AlaProProVal1ThrGluAla 188
Db 629 CCAAGGCCACGAGCGGAGCGAGCGCTCCCTCCGACCTCCCACTCCACC----- 679
QY 189 ProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProVal1GlyAlaThrAsn 208
Db 680 -----CCAACGACCCACGACGACTCTCTGTGCAACCAAGCTCTCTATGCGAGATGATCG 733
QY 209 ProVal1Val1ArgArgPheGlyThrAlaThrVal1AlaGlySerThrVal1ThrSerAsnGly 228
Db 734 GCT-----GCCGACACGACGAGTCCCATCGAGTGTCCACTTCTCGT 778
QY 229 MetTrpPheSerGlyArgAspGlyAspLeu1LeuAsn1AsnAlaGlyThrVal1Le 248

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 3, 2003, 21:28:33 ; Search time 146 Seconds

(without alignments)
3442.734 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670

Sequence: 1 MTVTAINSONQKPKIKLGL.....LFFFRISRNQVYVPLTLK 322

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1085931 segs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+D2n.model -DEV=xlh
-Q=/cgn2_1/USPFO.spool/US10018706/runat_30062003_091107_23899/app_query.fasta.1.519
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdd -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MALLEN=2000000000 -USER=US10018706@cgn_1.1.118@runat_30062003_091107_23899
-NCPU=6 -ICU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSEPLCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PC1_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PC1US_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	219.5	13.1	1830121 9	US-10-329-960-1 Sequence 1, Appli
2	209	12.5	2179 9	US-10-096-808-4 Sequence 4, Appli
3	187	11.2	9025608 9	US-10-156-761-1 Sequence 1, Appli
4	182.5	10.9	903 9	US-10-156-761-1629 Sequence 1629, Ap

5	175	10.5	609 9	US-10-156-761-4684 Sequence 4684, Ap
6	175	10.5	9025608 9	US-10-156-761-1 Sequence 1, Appli
7	153	9.2	3353 9	US-09-927-827-41 Sequence 41, Appli
8	146.5	8.8	615 9	US-09-738-626-2741 Sequence 2741, Ap
9	146.5	8.8	3309400 9	US-09-738-626-1 Sequence 1, Appli
10	145	8.7	949 9	US-10-092-243A-8 Sequence 8, Appli
11	136.5	8.2	1191 9	US-10-156-761-7229 Sequence 7229, Ap
12	135	8.1	1830121 9	US-10-329-960-1 Sequence 1, Appli
13	130.5	7.8	15872 9	US-09-860-846-1 Sequence 1, Appli
14	130.5	7.8	15872 9	US-09-868-384B-1 Sequence 1, Appli
15	130.5	7.8	15872 9	US-09-836-821-1 Sequence 1, Appli
16	130.5	7.7	15872 10	US-09-661-289-1 Sequence 4, Appli
17	128.5	7.6	1520 9	US-10-087-667-4 Sequence 165, App
18	127	7.6	7277 9	US-09-468-147-165 Sequence 165, App
19	127	7.6	7277 9	US-09-468-147-165 Sequence 165, App
20	126	7.5	774 9	US-10-156-761-2407 Sequence 2407, Ap
21	125.5	7.5	942 10	US-09-770-445-354 Sequence 354, App
22	125	7.5	903 9	US-09-468-147-141 Sequence 141, App
23	122.5	7.3	2436 9	US-10-063-547-99 Sequence 99, Appli
24	122.5	7.3	2436 9	US-10-174-590-309 Sequence 309, App
25	122.5	7.3	2436 9	US-10-176-758-309 Sequence 309, App
26	122.5	7.3	2436 9	US-10-063-616-99 Sequence 99, Appli
27	122.5	7.3	2436 9	US-10-175-37-309 Sequence 309, App
28	122.5	7.3	2436 9	US-10-063-502-99 Sequence 99, Appli
29	122.5	7.3	2436 9	US-10-173-706-309 Sequence 309, App
30	122.5	7.3	2436 9	US-10-175-738-309 Sequence 309, App
31	122.5	7.3	2436 9	US-10-175-752-309 Sequence 309, App
32	122.5	7.3	2436 9	US-10-176-482-309 Sequence 309, App
33	122.5	7.3	2436 9	US-10-176-757-309 Sequence 309, App
34	122.5	7.3	2436 9	US-10-176-913-309 Sequence 309, App
35	122.5	7.3	2436 9	US-10-180-552-309 Sequence 309, App
36	122.5	7.3	2436 9	US-10-180-557-309 Sequence 309, App
37	122.5	7.3	2436 9	US-10-173-700-309 Sequence 309, App
38	122.5	7.3	2436 9	US-10-174-572-309 Sequence 309, App
39	122.5	7.3	2436 9	US-10-174-579-309 Sequence 309, App
40	122.5	7.3	2436 9	US-10-174-582-309 Sequence 309, App
41	122.5	7.3	2436 9	US-10-174-588-309 Sequence 309, App
42	122.5	7.3	2436 9	US-10-175-739-309 Sequence 309, App
43	122.5	7.3	2436 9	US-10-175-740-309 Sequence 309, App
44	122.5	7.3	2436 9	US-10-175-743-309 Sequence 309, App
45	122.5	7.3	2436 9	US-10-176-488-309 Sequence 309, App

ALIGNMENTS

RESULT 1
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
GENERAL INFORMATION:
APPLICANT: Pletschmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm
FILE REFERENCE: PB186P1
CURRENT APPLICATION NUMBER: US/10/329, 960
PRIOR FILING DATE: 2003-01-02
PRIOR APPLICATION NUMBER: US 09/643, 990
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: US 08/487, 429
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426, 787
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)

```

Db      750082 CTTTCAGGTAAATATCATCCAAAGTCTTCACAGCAGATGGCGGTAAACAAGAAATTGAT 750141
Oy      231  PheSerGIYAgaSergLYAspleuILEasnaIsaSerAsnaIGLYThrValIleGlnAla 250
Db      750142 ATTATGGTGTACAGGAGCAAGCTGTAAAGGGCGTGACAGGGCGGAATAGTGTGCA 750201
Oy      251  AsphIsaMetLaspGLYAlaSer-----IleValIleGlnHisThrAsnGLYPheVal 268
Db      750202 GCGCAATGCTTTACCGTGGTTACGGTATTTAAATATATCATCAAAATATATGATGATTTT 750261
Oy      269  SerSerTYrIleHisIleIleYAspAlaGlnValIlysthrGLYAspThrValArgThrGLY 288
Db      750262 AGTCGCTATGGCACAATACGACAACAATTCCTGTGCGCGCATCAACAAGAAATCAAGACAGGT 750321
Oy      289  GlnArgIleAlaSerMetLYsaSngInProSerGLYAlaAlaIleuPheGluPheArgIle 308
Db      750322 CAACAACATCGCAAAATATGGTAGCTCTGTACAAATATACCGTGAACCTCCACTTTGAAATT 750381
Oy      309  SerArgAsnGLYValTY-ValAspProIeuThrValIleu 321
Db      750382 CGCTATAAAGGTAAATCATGAGTCCAGTCAAGTACTTA 750420

RESULT 2
US-10-098-808-4
; Sequence 4, Application US/10098808
; Publication No. US20030007981A1
; GENERAL INFORMATION:
; APPLICANT: Potier, Andrew A.
; APPLICANT: Rioux, Clement
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
; FILE REFERENCE: 9000-0049.20
; CURRENT APPLICATION NUMBER: 2002-03-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/405,728
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/267,749
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2179
; TYPE: DNA
; ORGANISM: Haemophilus somnus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (872)..(1906)
US-10-098-808-4

Alignment Scores:
Pred. No.:      1,21e-12      Length:      2179
Score:          209.00      Matches:      63
Percent Similarity: 44.09%      Conservative: 42
Best Local Similarity: 24.80%      Mismatches: 106
Query Match:      12.51%      Indels:      36
DB:              9          Gaps:      6

US-10-018-706-2 (1-322) x US-10-098-808-4 (1-2179)

Oy      73  TyrGlnValLYsGlnGLYAspThrValSerLYsIleAlaGlnArgTYrGLYleuAsnTP 92
Db      1229 TACAAGAATGCAAAAGCGGATACCATCTTCTTATGCTTATATTCACGCAATGATATA 1288
Oy      93  ArgGluIleGLYHisIleIleAsnAsnLeuAsnSerSerTYrThrIleTYrThrGLYGlnTP 112
Db      1289 AAAAGATTGGCCACATAAATATATATGCTGAGCAATCATCATGATTTGGACAAGTA 1348
Oy      113  LeuThrLeuTPSerGLY-----AspleuLYsValArgGLYAspSerIleSerSer 129
Db      1349 TTGAAAATATGCAAATATATATTCGCCGATAGCAATATGATATACCAACACAGCATTAATATGA 1408
Oy      130  GlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProVal 149

```

Db	1409	TCAGAGGTGACGACAAATACAGTCAATGAG---ACATGGAATGCTAAATAAACCAACAAAT	14655
Qy	150	GIingInhISProAlaValGImlnlyserProthProProValValValIlyellyProth	169
Db	1466	GACCAATGAAACCCGTTGCTACACCAACACATTCAACAATGCCATCAATATAA-----	1519
Qy	170	ProthProProValValGIingInProAlaProValAlaProProValThIgluaIaPro	189
Db	1520	---ACACTCCAGCCACCTCAATATACCTTGAGATTGGCCA-----	1558
Qy	190	PhealathnIySerSergIyValwclnPhearGTYrProValGIyAlaThIasnPro	209
Db	1559	-----ACAAATGGAAA	1570
Qy	210	ValValaIaGArGpHeGIYThAlaIaThValaIaGIserThValThIserSangIyMet	229
Db	1571	ATWTATTCAGAGATTTCACAGTGCATGAGAGC-----AATAAGGATAT	1615
Qy	230	TrpPheSerGIyAxAsgAspGIyAspLeuIleAsnIlaSerAsnAlaGIYThValIleGIln	249
Db	1616	GATATTACGGGTTCGTGGAAGAAGCTTTATGACAGCAGCGTGTGAGCAGAGTTGTAT	1675
Qy	250	AlaAspHisAsnMetAspGIyAlaIaSer-----IleValIleGIlnHisThIserGIyPhe	267
Db	1676	GCCGGAGACGCTTACGTCGATACGTATTAATTAATATATATATTAACATAAAGACAGTTAT	1735
Qy	268	ValSerSerTYrIleHisIleIyAspAlaGIlnValIySthrGIyAspThValaIaGthr	287
Db	1736	TTAAATGCTTATGCAATATGAATGAATGATCTCGTCAAAGATCAAGCAAGAAATTAAGCG	1795
Qy	288	GIyGlnaIaGIleAlaSerMetIyAsnGIlnProSergIyAlaIaIaLeuPheGIlnPhearG	307
Db	1796	GGTCAACAATAATGCTAAATATGGGAATCTTGGAACAAACACATCAACTCCATTCTGA	1855
Qy	308	IleSerArGAsnGIyValTYrValaAspProLeuThValIeu	321
Db	1856	ATTGCTTATTAAGTCAATCACTGATGATCCATGATCAATGATTTTA	1897

```

RESULT 3
US-10-156-761-1/c
Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OKURA, SATOSHI
APPLICANT: IKEDA, HARO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, YASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO: 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

Percent Similarity:	34.35%	Conservative:	36
Best Local Similarity:	24.38%	Mismatches:	117
Query Match:	11.20%	Indels:	120
DB:	9	Gaps:	16

QY	27	ThrcyVslleuAlaaglycysAlaSerlyProthrTryransSerThrSerGlySerGly	46
Db	2009653	ACAGCTCACTCGAGGCACTGAGAGAGAAATGACC-----ATGCCCGCAAGGT	20096603
QY	47	SerIlaIyIy-----ThSerGlySerGly	54
Db	2009602	AAGCAACCGCGCCCAAGTCCCGCGTTTACCCGCTCATCCGCTGCGGCGGAGCGGT	20095433
QY	55	GlyLeuAlaIle-----GlySerGlnValIleThrAspSer	66
Db	2009542	GGCCCGCGGCTCGCGCTCCGCTGATGGGGGCAACCGGCGCCAGCGCCGACCCCGACG	20094833
QY	67	GlnGlyValProAnaIy-----	72
Db	2009482	CGGCGCGTCTCGGAAAGGTCTCCGACGCCCGTGCSCCGGAGACGAGGCGCGCGAG	20094233
QY	73	-----TyrGlnValIySglnGlyAspThrValSerIyIleAlaGln	86
Db	2009422	AAGACGGGACAGAAAGCATATGCGGTACGGCGGGGTATCTCCCTCGAAGATGCCGAC	20093633
QY	87	ArgTyrGlyLeuAnaIyPArgGlnIleGlyHisIleAnaIyAnaIySerSerTyrThr	106
Db	2009362	GAGCAGAGCGTC-----	2009351
QY	107	IleTyrThrGlyGlnIlePLeuThrLeuTrpSerGlyAspLeuIyValArgGluArgSer	126
Db	2009350	-----ACCGCGGCTGGAAGAAAGCTCACTCGGACAAAC-----CGCTCC	2009312
QY	127	IleSerSerGlyValAnaIyIle-----AlaHis	135
Db	2009311	GCCATCGGCGGCGACCCGACGGTATCACCCCGGTGTAAGTCAGCATCGGCGCCAG	2009252
QY	136	ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal	155
Db	2009251	TCCGCGTGAAGCGCGCCAGCAAGTCTGTC-----	2009222
QY	156	GlnIyPProThrProProValValValIleIyIySerProThrProThrProProValVal	175
Db	2009221	-----ACGGCAGACGAGCGCGCAACGGGGGTAAATGCCCCACG-----GCG	2009180
QY	176	GlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySerSer	195
Db	2009179	AAGACCCCGCGTTCGAAAGACACACACGCGCCACCGGCGCGACACACACACCGCGCGC	2009120
QY	196	GlyValMetGlnPheArgTyrProValGlyAlaIleThrAsnProValValArgArgPheGly	215
Db	2009119	GGC-----ThACCCCTCCGCGTTCAGCGC-----	2009096
QY	216	ThrIleAlaThrValAlaGlySerThrValThrSerAnaIyMetTrpPheSerGly-----	233
Db	2009095	-----GCCACCATCGTACCGCTACAGCCGCGGCGAGCATGTGTCTCAAGGGGTACCAC	2009039
QY	234	-----ArgAspGlyAspLeuIleAsnAlaSerAnaIyGlyThr	246
Db	2009038	ACGGCGCTCGACTTCGTGCTCCGACCGGCAACCACTCAAGGCGTGCAGCGCGGCGACC	2008979
QY	247	ValIleGlnAlaAsp-----HisAsnMetAspGlyAlaSerIleValIleGlnHisThrAsn	265
Db	2008978	GTCTCTTCGCGCGGCTGCGGCGCGCGCGCTACGCGACACAGCATGCTGCTGCGGACGCCAT	2008919
QY	266	GlyPheValSerSerTyrIleHisIleIyIyAspAlaGlnValIyThrGlyAspThrVal	285
Db	2008918	GGCGCATCTACGCAATGACGCCCATGTCTCCAGTCTCCGCTTCCACGCGGCGAGAGGTG	2008859
QY	286	ArgThrGlyGlnArg-----IleAlaSerMetIyAsnGlnProSerGlyAlaAlaIleu	303

Db 2008858 GCCAGGCGCGCAGCTGGGCTCTGTGTGCGACCGGCAATGTCACGGAGCCGACCTG 2008799
Qy 304 ---PheGluPheArgIleSerArgAsn---GlyValTyrValAspProLeuThrValLeu 321
Db 2008798 CACTTCGAGATCCGTAGACCCCTCTTACGCTCGACGTGAGCCGCGCTTACTTG 2008739
Qy 322 Lys 322
Db 2008738 CGC 2008736
RESULT 4
US-10-156-761-1629
Sequence 1629, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1629
LENGTH: 903
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(903)
US-10-156-761-1629
Alignment Scores:
Pred. No.: 2,52e-10 Length: 903
Score: 182.50 Matches: 83
Percent Similarity: 34.50% Conservative: 35
Best Local Similarity: 24.27% Mismatches: 107
Query Match: 10.93% Indels: 117
DB: 9 Gaps: 15
US-10-018-706-2 (1-322) x US-10-156-761-1629 (1-903)
Qy 46 GlySerHisArg-----ThSerGlySer 53
Db 13 GGTAGACACCGCCGCCAGGTCCCGCGTTTACCGCGTGCATGCGCGCGGAAC 72
Qy 54 GlyGlyLeuAlaIle-----GlySerGlnValIleThrAsp 65
Db 73 GGTGGCGCGCGCTCGCGCTCCGCTGATGGGGCGAGACCGCGCCACCGCCAC 132
Qy 66 SerGlnGlyValProAsnArg----- 72
Db 133 ACGCGCGCGCTCCCGGAAGGTCTCCGACACCCCGTTCGCGGAGACGAGGCGCGCC 192
Qy 73 -----TyrGlnValysGlnGlyAspThrValSerIleAla 85
Db 193 GAGAGAGCGCGACGAGACGTATGCGGTGCGGGGTGACTCCCTTCGAAAGATCGCC 252
Qy 86 GlnArgTyrGlyLeuAsnTyrArgGlnIleGlyHisIleAsnAsnLeuAsnSerSerTyr 105
Db 253 GACGAGCAGACGCTC----- 267
Qy 106 ThrIleTyrThrGlyGlnTyrLeuThrLeuTyrSerGlyAspLeuLysValArgGlnArg 125
Db 268 -----ACCGGCGCGCTGGAAGAGCTTACTCGACAC-----CGC 303

Qy 126 SerIleSerSerGlyValAsnThr-----Ala 134
Db 304 TCCGCACATCGCGCGCGACCCGACCGCTATCCACCCCGGTGTGAGCTGACATGCGGGCC 363
Qy 135 HisThrProSerProValAlaValGlnSerSerArgProProValGlnHisProAla 154
Db 364 AAGTCCGCTCGACGCGCGCCACGCACTCGTCC----- 396
Qy 155 ValGlnIleProThrProProValValValValIleValIleValIleProThrProProVal 174
Db 397 -----ACGGCGAGAGAGCGCGCCACGCGGTGAGTCCGCCACG----- 435
Qy 175 ValGlnIleProAlaProValAlaProProValThrGlnIleAlaProPheAlaThrGlySer 194
Db 436 GCGAAGACCCCGCGGTGAGAGACACACGCGCCACCGCGCGCGGACACACACCGCGC 495
Qy 195 SerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValValArgAspPhe 214
Db 496 GCCGCG-----TACACCCCTGCGGTGACGCG----- 522
Qy 215 GlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTyrPheSerGly---- 233
Db 523 -----GCCACATCGGTACCGCTTACAGACCGCGCGACATGTGTCCAGCGGTAC 576
Qy 234 -----ArgAspGlyAspLeuIleAsnAlaSerAsnAlaGly 245
Db 577 CACACCGCGCTGCACTTGTGTGTCGCCACCGCGCACCACTCAAGCGTGCAGCGCGCGC 636
Qy 246 ThrValIleGlnAlaAsp---HisAsnMetAspGlyAlaSerIleValIleGlnHisThr 264
Db 637 ACGGTGCTCGCGCGCGCTGCGGGCGCGCGTACGACAGAGTGCATGTCGCGCACGCC 696
Qy 265 AsnGlyPheValSerSerTyrIleHisIleIleAspAlaGlnValIleThrGlyAspThr 284
Db 697 GATGCGCAGTACTCGACGTACGCGCGCCACATGTCCAGCTTCCTCCACGCGGCGACAGC 756
Qy 285 ValArgThrGlyGlnArg-----IleAlaSerMetLysAsnGlnProSerGlyAlaAla 302
Db 757 GTGCGCGAGGCGCGCGACGCTGCGCTCTGTGTCGACCGGCAATGTCACGGAGCCGAC 816
Qy 303 Leu---PheGluPheArgIleSerArgAsn---GlyValTyrValAspProLeuThrVal 320
Db 817 CTGCACTTCGAGATCCGTAGACCCCTCTTACGCGTTCGAGCGTGCAGCCGCGTCTTAC 876
Qy 321 LeuLys 322
Db 877 CTGCGC 882
RESULT 5
US-10-156-761-4684
Sequence 4684, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 4684
LENGTH: 609
TYPE: DNA


```

US-09-927-827-41
; Sequence 41, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramsdell, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 41
; LENGTH: 3353
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2353)
US-09-927-827-41

Alignment Scores:
Pred. No.: 2,366-06 Length: 3353
Score: 153.00 Matches: 61
Percent Similarity: 42.47% Conservative: 32
Best Local Similarity: 27.85% Mismatches: 85
Query Match: 9.16% Indels: 42
Gaps: 11

US-10-018-706-2 (1-322) x US-09-927-827-41 (1-3353)

QY 111 GINTPLeuThrLeuTyrPserGlyAspLeuValAlaGlnArgSerIleSerSergly 130
   ::::: :::::
Db 46 CGCTGGAGCAGTTGCTGGCCAACTCTGCGTGCAGCCGCGCCAGAGCCG----- 93

QY 131 ValaenThAlaHisThrProSerProValAlaValGlnSerSerArgProPro--Val 149
   94 -----ACGCTGAGCGCGCGCGCGCGCCCAAGCGTGGCGCGCGCGCAAGCCG 138

QY 150 GInGlnHisProAlaValAlaGlnIysProThrProProValValValValIyIysProThr 169
   ::::: :::::
Db 139 CGGCGCAGGCCAAGCGCGCGCAAAACCGATGCG-CCG-----GACGCGCCCGGC 185

QY 170 ProThrProProValValAlaGlnGlnProAlaProValAlaProProValThrgIuaPro 189
   ||||| |||||
Db 186 AAGACTCGCGCCCAAGGTCGTGGCCAGACGACCG-----GGGCCA 224

QY 190 PheaIaThrGlySerSerGlyValMetGlnPheArgTyrProValGlnAlaThrAsnPro 209
   ::::: :::::
Db 225 AAGTCGTGTGGCTGTAGT-----TGGCGGTGTTCGGGCAAT----- 260

QY 210 ValValAlaArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
   ::::: :::::
Db 261 CTGCTGGCGCGCTTCAACGCCACAGCTGCCGAGTGGCAAT-----ACAGCAAGGGCGTG 314

QY 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGln 249
   ||||| |||||
Db 315 CTGATCGCGCGCGCCCAAGGGCAGACAGCGATCCGCGGTGGCCGATGGCAAGCGTGT 374

QY 250 AlaAspHisAsnMetAsp---GlyAlaSerIleValIleGlnHisThrAsnGlyPheVal 268
   ::::: :::::
Db 375 TCCGACTGTGATGACCGGCTACGGCAGATGATTCGTGATTCGATTCGCGCATGGCTCATAG 434

QY 269 SerSerTyrIleHis-----IleYAspAlaGlnValIySthrgIyAspThr 284
   ||||| |||||
Db 435 AGCCTGTACGGCACAACAGACACTTGTCTGCGGATGCGC-----GGCGCCAGC 482

QY 285 ValArgThrGlyGlnArgIleAlaSerMet---LysAsnGlnProSerGlyAlaAlaLeu 303
   ::::: :::::
Db 483 ATCAACAACGTGGAGAACCGGTGGCCAAAGTCGATTCGGAGCGGCGGACAGGTGTGCGGGCG 542

QY 304 PheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeuThrValIeulys 322

```

[illegible]

QY 271 TTTTLEHISLELYASAPALAGINValValThrglyAspThrValArgThrGlyGlnArg 290
DB 436 TACGACACATGGAAATACCTTCTACGTCCTCGCGCGACCGCTCGACAGAGCGCAGAA 495
QY 291 TLEALASerMetLYAspGlnPro--SerGlyAlaAlaLeuPheGluPheArgIleSer 309
DB 496 ATCGAGGAATGGGACGCCAAGATTCTCCACCGGCTCCACCTTCACCTTCGAGATTCAC 555
QY 310 ArgAsnGlyValIYr--ValAspPro 317
DB 556 CCAGACGGCGTCAACCCCGATCGACCCA 582
RESULT 9
US-09-738-626-1
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
PRIORITY FILING DATE: 2000-12-28
PRIORITY FILING DATE: 1999-12-16
PRIORITY FILING DATE: 1999-12-16
PRIORITY FILING DATE: 1999-12-16
PRIORITY FILING DATE: 2000-04-07
PRIORITY FILING DATE: 2000-04-07
PRIORITY FILING DATE: 2000-04-07
PRIORITY FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1
Alignment Scores:
Pred. No.: 0.182 Length: 3309400
Score: 146.50 Matches: 64
Percent Similarity: 44.44% Conservative: 20
Best Local Similarity: 33.86% Mismatches: 29
Query Match: 8.77% Indels: 39
Gaps: 10
US-10-018-706-2 (1-322) x US-09-738-626-1 (1-3309400)
QY 137 ProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
DB 2649656 CCGCACCGCTACCG-----CCACCGACGACACGACATCG-----CGC 2649694
QY 157 LysProThr-----ProProValAlaValAlaValAlaValAlaValAlaValAla 172
DB 2649695 AAACCGACTACGACGCGCTTCTCCCGCGCGTGG-----CCGACACCGTGCAG 2649742
QY 173 ProValValGln--GlnProAlaProValAlaProProValThrGlnAlaProPheAla 191
DB 2649743 AACCTGACGAGATGCGACACCGCGCTC-GCACCGACCGCCCGGATGCGCGCCACGCA 2649801
QY 192 ThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValAla 211
DB 2649802 AACGGCAC-----TTCACTCGAGATTCGGA 2649828
QY 212 ArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPhe 231

DB 2649829 CCACGTTGGGAGAAC-----TTCCACACCGCATTCACATC 2649864
QY 232 SerGlyArgAspGlyAspLeuIleAsnAlaSerAlaGlyThrValIleGlnAlaAsp 251
DB 2649865 GCAAATCTCAATCGGACCCCAATCTACCGCTGATGCGCGGACGATCTGACCTGCGC 2649924
QY 252 HisAsnMetAsp--GlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSer 270
DB 2649925 CCAGATCCGGCTATGACAGTGCATTCGACGACGACGACGATCCATCTTCATC 2649984
QY 271 TTTTLEHISLELYASAPALAGINValValThrglyAspThrValArgThrGlyGlnArg 290
DB 2649985 TACGACACATGGAAATACCTTCTACGTCCTCGCGCGACCGCTCGACAGAGCGCAGAA 2650044
QY 291 TLEALASerMetLYAspGlnPro--SerGlyAlaAlaLeuPheGluPheArgIleSer 309
DB 2650045 ATCGAGGAATGGGACGCCAAGATTCTCCACCGGCTCCACCTTCACCTTCGAGATTCAC 2650104
QY 310 ArgAsnGlyValIYr--ValAspPro 317
DB 2650105 CCAGACGGCGTCAACCCCGATCGACCCA 2650131
RESULT 10
US-10-092-243A-8/c
Sequence 8, Application US/10092243A
Publication No. US20020197625A1
GENERAL INFORMATION:
APPLICANT: Hillman, Jeffrey D.
TITLE OF INVENTION: Microbial polynucleotides Expressed During Infection of
FILE REFERENCE: MEH800-5051
CURRENT APPLICATION NUMBER: US/10/092, 243A
PRIORITY FILING DATE: 2002-05-21
PRIORITY FILING DATE: 1999-08-06
PRIORITY FILING DATE: 1999-08-06
PRIORITY FILING DATE: 2000-08-04
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 8
LENGTH: 949
TYPE: DNA
ORGANISM: Actinobacillus actinomycetemcomitans
FEATURE:
NAME/KEY: misc feature
LOCATION: (538)
OTHER INFORMATION: N stands for any nucleotide.
US-10-092-243A-8
Alignment Scores:
Pred. No.: 2.93e-06 Length: 949
Score: 145.00 Matches: 57
Percent Similarity: 37.04% Conservative: 23
Best Local Similarity: 26.39% Mismatches: 62
Query Match: 8.68% Indels: 74
Gaps: 9
US-10-018-706-2 (1-322) x US-10-092-243A-8 (1-949)
QY 73 TTTGlnValValGlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrp 92
DB 537 TATACCGTACGAAAGGTGACACCACTGTCTTATCGCTTACATTTCTGTGTTGGATGTG 478
QY 93 ArgGlnIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112
DB 477 AAAGATATGGCGCGCTTAAATATATGTGCCAATCCTTACACTTAAAGGTGGAACCAAC 418
QY 113 LeuThrLeuTrpSerGlyAspLeuIleValArgGlnArgSerIleSerSerGlyValAsn 132
DB 417 TTA-----AAGTGAACCGCGCGC----- 400
QY 133 ThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152

Db 399 -----ACACGGTTACGGTGCACAAAGCTGACCCCAATGCA--- 358
 Qy 153 ProAlaValGlnLysProThrProProValValValLysProThrProThrPro 172
 Db 357 -----CCACGGCTCACCCAA 343
 Qy 173 ProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThr 192
 Db 342 CCGGTTACGCAAGGTGGTGAACCGAC---GTAAGTTACACCCCGGCGCAAC 292
 Qy 193 GlySerSerGlyValMetGlnPheArgGlyProValGlyAlaThrAsnProValAlaArg 212
 Db 291 GGCACA-----CAATATGTTCCGACGGCAGATTAACCGCCCATTAAGCC 244
 Qy 213 ArgPheGlyThrAla-----ThrValAlaGlySerThr--- 223
 Db 243 GCGGTAGGTACTGTCGACCGGCAACCAATGACCGATTAACCGGCGACCGCACCG 184
 Qy 224 ValThrSerArgLysMetTrp-----PheSer 232
 Db 183 GCAATTCACACGTCGACGCGCTGCGCCGACCAAGCAATGTCATTCAAGTTCTCT 124
 Qy 233 GlyArgAspGly-----AspLeuIleAsn 240
 Db 123 AACGCGACGCGGTAAACAAAGCATGACATGAGTGTCTCCGACGACAAAGCATCAT 64
 Qy 241 AlAserAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGly 256
 Db 63 GCCCGCGCGCAGGTGAGTGTTCACGCGGTAAACGCTTACGCGGT 16

RESULT 11

US-10-761-7229
 ; Sequence 7229, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT FILING DATE: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 7229
 ; LENGTH: 1191
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1191)
 ; US-10-156-761-7229

Alignment Scores:

Pred. No.: 3.31e-05 Length: 1191
 Score: 136.50 Matches: 82
 Percent Similarity: 36.70% Conservative: 27
 Best Local Similarity: 27.61% Mismatches: 126
 Query Match: 8.17% Indels: 63
 DB: 9 Gaps: 15

US-10-018-706-2 (1-322) x US-10-156-761-7229 (1-1191)

Qy 43 SerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySer----- 60
 |||||
 :|||

Db 421 TCCGCGACGCGACACACACGCGGACCGGAGCACCAAGCGGACCGTCCGTGATGTGAAG 480
 Qy 61 ---GlnValIleThrAspSerGln---GlyValProAsnArgGlyGlnValLysGlnGly 78
 Db 481 CCACAGACGACACCGCATGTCGACAGCGGCGACCGCCGAGATGACACCGCTTCGACGCG 540
 Qy 79 AspThrValSerLysIleAlaGlnArgGlyGlyLeuAsnTrpArgGlnIleGlyHisIle 98
 Db 541 GACACGCTCTCCGGATCGCC-----GACTCCGCGACAGTC 576
 Qy 99 AsnAsn-LeuAsnSerSerGlyThrIleGlyThrGlyGlnTrpLeuThrLeuTrpSerGly 118
 Db 577 CCGGCGGTTGGGAGCAGCTTACGCGCCGACCGCGG-----GGACC 618
 Qy 118 YAspLeuLysValArgLysArgSerIle---SerSerGlyValAsnThrAlaHisIlePr 137
 Db 619 ATCGCGCCGACCGCGACCTGATCTTCCGGCGACGCGGTGATCTGCGGCGCAAGCGC 678
 Qy 137 oSerProValAlaValAlaInsSerSerArgProProValGlnGlnHisProAlaValGlnLys 157
 Db 679 ACGACCGGACCGACGACGCGCTCCACGTCGACCCCAAGTCGT----- 721
 Qy 157 sProThrProProValValValValLysLysProThrProThrProProValValGlnGly 177
 Db 722 -----CGACGCTCTCTCAAGTCTCGACGCGAAGAGCG 756
 Qy 177 nProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyVal 197
 Db 757 TCCTCGG---ACAGACCG-----AGAGCGGACGACGAGCCCTCCCTCGTGGC--- 804
 Qy 197 lMetGlnPheArgGlyThrValIleGlnAlaThrAsnProValValArgArgPheGlyThrAl 217
 Db 805 -----CCGCTGAACGCC-----GGACGCG 824
 Qy 217 aThr-----ValAlaGlySerThrValThrSer-----AsnGlyMetTrpPh 231
 Db 825 GACGCAATACCATCGCGCGGTTGCTCGGTGCGAAGGCGTCATCACACCGGCGTCGACTT 884
 Qy 231 eSerGlyArgAspGlyLysPheLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAs 251
 Db 885 CCGCGTCCACACGCGGACCTCCGTCGAGAGCGGTGCGCGGCGGCGACGTCGTCACGCGCG 944
 Qy 251 pHisAsnMetAsp---GlyAlaSerIleValIleGlnHisIleThrAsnIlePheValSerSe 270
 Db 945 CTGGGCGGCGCTCTTCGCTTACCAAGTGTGCTCCGCGACCGGCGGCTTACCGCA 1004
 Qy 270 rTyrlleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArg 290
 Db 1005 GTACGCCCATCTGTGCGCATCTCCGTGAAGGCGGCGACGCGTGAACGCGGCGACGCG 1064
 Qy 290 gIle-----AlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheGlyLysPheArg 307
 Db 1065 TATCGCGCGCTCGGCGGTCCACAGGCGCAACAGACGCGGCGCGATTCGACTTCGAGTGGC 1124
 Qy 307 gIleSerArgAsn---GlyValTyValAspProLeuThrValLeuLys 322
 Db 1125 GACGCGGCGCGGCTTCCGACGTGACGTGATCCGCTTATGTGAGG 1173

RESULT 12

US-10-329-960-1/c
 ; Sequence 1, Application US/10329960
 ; Publication No. US20030099277A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fleischmann et al.
 ; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragment
 ; FILE REFERENCE: P186P1
 ; CURRENT FILING DATE: US/10/329,960
 ; PRIOR FILING DATE: 2003-01-02
 ; PRIOR APPLICATION NUMBER: US 09/643,990
 ; PRIOR FILING DATE: 2000-08-23
 ; PRIOR APPLICATION NUMBER: US 08/487,429
 ; PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: US 08/426,787
PRIOR FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29258)..(29258)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45393)..(45393)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119824)..(119824)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

```
LOCATION: (131360)..(131360)
FEATURE:
OTHER INFORMATION: n equals a, t, g or c
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 1.37 Length: 1830121
Score: 135.00 Matches: 63
Percent Similarity: 39.29% Conservative: 36
Best Local Similarity: 25.00% Mismatches: 111
Query Match: 8.08% Indels: 42
DB: Gaps: 10

US-10-018-706-2 (1-322) x US-10-329-960-1 (1-1830121)
QY 74 GlnVallysgInglyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrpArg 93
Db 820115 CAAAAAACAACAACAAGCATTCGAAAAACACAGCAGCATCAA----- 820068
QY 94 GluIleGlyHisIleAsnAsnLeuAsnSerSeryTrpIleTyrThrGlyGlnTrpLeu 113
Db 820067 -----TCTACGCTGGAATGAATCAATTAATTTAGCCCTGATCAAGATTAATTTGAAT 820014
QY 114 ThrLeuTrpSerGlyAspLeuIleValArgGluArgSerIleSerSerglyValAsnThr 133
Db 820013 ACA-----CTAAAGCAACAACAGACACTTCGCAAGAAATTCACGA 819969
QY 134 AlaHisThrProSerProValAlaValGlnSerSeraTrpProProVal---GlnGlnHis 152
Db 819968 GCTGAACAAGCAGCGCGCAACAAGAAACGTCGAAGAAGAGCAGCTTCCTCAAGCCAA 819909
QY 153 ProIleValGlnIleProThrProProValValValValValValValValValValVal 172
Db 819908 AAGGCTGAAGAAACAACGAACA----- 819888
QY 173 ProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThr 192
```

```
Db 819887 -----TCAAAACCTTATCAACCAACTGTGCAAGACGCCAATTACTTAAT 819843
QY 193 GlySerSerglyVal-----MetGlnPheArgTyrProValGlyAlaThrAsn 208
Db 819842 AGTACAGCGGTTATGGGCGCGCAAAAAACAATATCTTACCAAGTTCTGCTTCA--- 819786
QY 209 ProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGly 228
Db 819785 ---ATTTCGATCACTTTTGGT-----TCTATCCACAGCGCAAGTACGTTGAAGGT 819735
QY 229 MetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle 248
Db 819734 ATGGTAATTCGCGCATCAGCAGGCACGCTGTAAACCAATTCGCTGCGCAGCGCATTT 819675
QY 249 GlnAlaAspHis---AsnMetAspGlyAlaSerIleValIleGlnHisThrAsnGlyPhe 267
Db 819674 TTACGGGGAATTTAAATGGTTATGTATATGTATTTGTTAAACAAGCGCAAACTGAT 819615
QY 268 ValSerSeryTrpIleHisIleLysAspAlaGlnValIleThrGlyAspThrValArgThr 287
Db 819614 TTAACTTATATGTGCTTCATCAAGCTGTATCACTGAATTTGTCAGCTTTTCAGCA 819555
QY 288 GlyGlnArgIleAlaSerMetLysAsn-----GlnProSerglyAlaAlaLeuPheGlu 305
Db 819554 GGGCAGGTTATTTGCTCAAGTAGAATAACAGGGGAATATACAGCTTCGCGCTTAT--- 819498
QY 306 PheArgIleSerArgAsnGlyValTyrValAspPro 317
Db 819497 TTGGTATTTAGCCGTAAAGAGACGCCAGTAATCCT 819462

RESULT 13
US-09-860-846-1/c
Sequence 1, Application US/09860846
Patent No. US20020164742A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/860,846
PRIORITY FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae

US-09-860-846-1

Alignment Scores:
Pred. No.: 0.00547 Length: 15872
Score: 130.50 Matches: 81
Percent Similarity: 38.75% Conservative: 43
Best Local Similarity: 25.31% Mismatches: 115
Query Match: 7.81% Indels: 81
DB: Gaps: 18

US-10-018-706-2 (1-322) x US-09-860-846-1 (1-15872)
QY 34 AlaSerIleAspThrTyrAsnSerThrSerglySerglySerHisArgThrSerglySer 53
Db 3195 GCGTCGCGCC-----TCGGCAGCAGCGCGCCACAGAGTCTCGGGGAG 3151
QY 54 -----GlyGlyLeuAlaIleGlySerglyValIle 63
Db 3150 GTGACGCGCGCGGAGTACGCGCAGCCATCCAGTATCGCATGCGCGCGGTGAG 3091
QY 64 ThrAsp-----SergInglyValProAsnArgTyrGlnValIleSergInglyAspThrVal 81
```

```

Db      3090 TCCGCCGGGGTCCGGGGGATATCCCGGCGGATCCGCTCCGGCGCGCGCGCGG 3031
Qy      SerLys-----IleAlaGlnArgTyrGlyLeuAsnTrpArglu----- 94
Db      3030 AGAGAGTCCCGCCAGTGGGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2971
Qy      95 ILeGlyHisIleAsnAsnLeuAsnSerSerTyrTrpIleTyrThrGlyIntPrLeuThr 114
Db      2970 CTGGGCGAGCGCGAGTCCCGCTGCTCGACGAGCGCGCTCCGCGAGC----- 2926
Qy      115 LeuTrpSerGlyAsp-----LeuIleValArgGluArgSer 126
Db      2925 -----TCGACGAGCATGAGGAGTCCAGCCAGCTCTTGAAGGTGAGCGCGAGTTCG 2872
Qy      127 IleSerSerGly-----ValAsnThrAlaHisTrpProSerPro 139
Db      2871 ACCGGGTCGGGTCGTCTGATCTGAGAGGACGGCGCGACGTCGTGGCGACGCGCTGTC 2812
Qy      140 ValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThr 159
Db      2811 GTCTCGGCGAGCGCGCGCGCATCACC-----CCACCGCAGCCCGCGCGCTTCG 2764
Qy      160 ProProValValValValLysLysProThrProThrProProValValGlnGlnProAla 179
Db      2763 CCC-----TCGCTTCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCG 2722
Qy      180 ProValAlaProProValThrGlnAlaProPheAlaThrGlySer----- 194
Db      2721 CCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2662
Qy      195 ---SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProValVal 211
Db      2661 CCGTCGACCAAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2602
Qy      212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
Db      2601 CCGGTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2542
Qy      226 Ser-----AsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db      2541 AGTGGCGGAGCAGCAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2497
Qy      244 AlaGlyThrValIleGlnAlaAsp---HisAsnMetAspGlyLysSerIleValIleGln 262
Db      2496 GCGGAGACCGCGCGCGCGCGCGCTCTGTCGTGCGGTACGAGCGCGCGCGCGCGAG 2437
Qy      263 HisThrAsnGlyPheValSerSerTyrIleHisIleLysAsp---AlaGlnValLysThr 281
Db      2436 ACCCGCTCGGAGCGAGACTCCAGAGAGGTCCGCGCGCGCGCGCGCGCGCGCG 2377
Qy      282 GlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 301
Db      2376 GCGTCGAGGAAGCTACGGGCTGCGG-----ACCTGTCACCCAGTACTCGGAGTGC 2323

```

```

; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

Alignment Scores:
Pred. No.: 0.00547 Length: 15872
Score: 130.50 Matches: 81
Percent Similarity: 38.75% Conservative: 43
Best Local Similarity: 25.31% Mismatches: 115
Query Match: 7.81% Indels: 81
DB: 9 Gaps: 18

US-10-018-706-2 (1-322) x US-09-988-384B-1 (1-15872)

Qy      34 AlaSerLysProThrTyrAsnSerThrSerGlySerHisArgThrSerGlySer 53
Db      3195 GCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3151
Qy      54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
Db      3150 GTGACCGCGCGCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3091
Qy      64 ThrAsp-----SerGlnGlyValProAsnArgTyrGlnValLysGlnLysAspThrVal 81
Db      3090 TCCGCCGGGGTCCGGGGGATATCCCGGCGGATCCCGGCGCGCGCGCGCG 3031
Qy      82 SerLys-----IleAlaGlnArgTyrGlyLeuAsnTrpArglu----- 94
Db      3030 AGAGAGTCCCGCCAGTGGGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2971
Qy      95 ILeGlyHisIleAsnAsnLeuAsnSerSerTyrTrpIleTyrThrGlyIntPrLeuThr 114
Db      2970 CTGGGCGAGCGCGAGTCCCGCTGCTCGACGAGCGCGCTCCGCGAGC----- 2926
Qy      115 LeuTrpSerGlyAsp-----LeuIleValArgGluArgSer 126
Db      2925 -----TCGACGAGCATGAGGAGTCCAGCCAGCTCTTGAAGGTGAGCGCGAGTTCG 2872
Qy      127 IleSerSerGly-----ValAsnThrAlaHisTrpProSerPro 139
Db      2871 ACCGGGTCGGGTCGTCTGATCTGAGAGGACGGCGCGACGTCGTGGCGACGCGCTGTC 2812
Qy      140 ValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThr 159
Db      2811 GTCTCGGCGAGCGCGCGCGCATCACC-----CCACCGCAGCCCGCGCGCTTCG 2764
Qy      160 ProProValValValValLysLysProThrProThrProProValValGlnGlnProAla 179
Db      2763 CCC-----TCGCTTCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 2722
Qy      180 ProValAlaProProValThrGlnAlaProPheAlaThrGlySer----- 194
Db      2721 CCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2662
Qy      195 ---SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProValVal 211
Db      2661 CCGTCGACCAAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2602
Qy      212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
Db      2601 CCGGTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2542
Qy      226 Ser-----AsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db      2541 AGTGGCGGAGCAGCAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2497
Qy      244 AlaGlyThrValIleGlnAlaAsp---HisAsnMetAspGlyLysSerIleValIleGln 262
Db      2496 GCGGAGACCGCGCGCGCGCGCTCTGTCGTGCGGTACGAGCGCGCGCGCGAG 2437
Qy      263 HisThrAsnGlyPheValSerSerTyrIleHisIleLysAsp---AlaGlnValLysThr 281

```

```

Db      2436 ACCCGGCGGAGCGAGCTCCAGGAAGGTCTCGCGCGCCGATTCCTCCAGGGTGCAGC 2377
Qy      282 GtAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 301
Db      2376 GCGTCGAGGAAGGCTACGGGCGCTGCGG-----ACCTGTCACACCGATCTCGGATCG 2323

RESULT 15
US-09-836-821-1/c
/ Sequence 1, Application US/09836821
/ Publication No. US2003087405A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/836,821
/ PRIOR FILING DATE: 2001-04-17
/ PRIOR APPLICATION NUMBER: 09/105,537
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 1
/ LENGTH: 15872
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-836-821-1

Alignment Scores:
Pred. No.: 0.00547 Length: 15872
Score: 130.50 Matches: 81
Percent Similarity: 38.75% Conservative: 43
Best Local Similarity: 25.31% Mismatches: 115
Query Match: 7.81% Indels: 81
Gaps: 18

US-10-018-706-2 (1-322) x US-09-836-821-1 (1-15872)
Qy      34 AlaSerLysProThrTyraSerThrSerGlySerHisArgThrSerGlySer 53
Db      3195 GCGTCGCGCGCC-----TCGGCAGCAGCGCCGACAGGTCTCGGGGAG 3151
Qy      54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
Db      3150 GTGACCGCGCGGGGTAGCGGAGCCGATGCCATGATTCGCGATGCGGTGCGGCGGTG 3091
Qy      64 ThrAsp-----SerGlnGlyValProAsnArgTyrglnValLysGlnIleAspThrVal 81
Db      3090 TCGCGCGGGGTCCGGCGGATTCCTCGTCCGCGGATCCGCTTCGCGCGCGGTG 3031
Qy      82 SerLys-----IleAlaGlnArgTyrglyLeuAsnTrpArgIle 94
Db      3030 AGCAGGTCCGCCAGGTGGCGGCGAGGCGCGCGCTCGGTGTCGAGAGACAGTCCG 2971
Qy      95 IleGlyHisIleAsnLeuAsnLeuAsnSerGlyThrIleTyrglyGlnTrpLeuThr 114
Db      2970 CTGGGCGAGCGCGCGCTCGGTGTCGAGCGCGCTTCGCGAGC-----2926
Qy      115 LeuTrpSerGlyAsp-----LeuLysValAlaArgIleArgSer 126
Db      2925 -----TCGACGAGCATGAGGAGTGAAGCCGATCCTTGAAGTGAAGCCGAGTTCG 2872
Qy      127 IleSerSerGly-----ValAsnThrAlaHisThrProSerPro 139
Db      2871 ACCCGGTCGGGTGCTGCTACTGACGAGACGCGCGCATGCTGCGCGACGCGCTCGTGC 2812
Qy      140 ValAlaValGlnSerSerArgProProValGlnIleHisProAlaValGlnLysProThr 159
Db      2811 GTCTCGTGGCGAGCCGCGCATGACCG-----CCACCGCGACCCGCGCTCTG 2764
Qy      160 ProProValValValValLysLysProThrProThrProProValValGlnIleProAla 179

```

```

Db      2763 CCC-----TCGCTCGCGCTCGCGCCGACGTCACACCGCGCGCGG 2722
Qy      180 ProValAlaProProValThrGlnAlaProPheAlaThrGlySer-----194
Db      2721 CCGGTCCCGCACCGCGGCTTCGCGTACGCGCGCGCGCGCGCGCGCTTCGCGCGCG 2662
Qy      195 ---SerGlyValMetGlnPheArgTyPro-----ValGlyAlaThrAsnProVal 211
Db      2661 CCGTCCGACCAAGTGGCGCTTCGCGTGAAGCGTGAAGTCCGACAGGGACCCCTACCTG 2602
Qy      212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
Db      2601 CCGGTGCTCCCGTGGCGGCGGCTCGCATGACAGTCGTCGCGCGCGAGAGAGCGTGTG 2542
Qy      226 Ser-----AsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAlaSerAsn 243
Db      2541 AGTCCGCGGAGCAGCGATCGGCGCTCGCGCGCGCC-----TTGCGCAGG 2497
Qy      244 AlaGlyThrValIleGlnAlaAsp---HisAsnMetAspGlyAlaSerIleValIleGln 262
Db      2496 GCGGAGACCGCGCTGGCGGCTCTCTGTCGCGTACGAGAGTCCGCGCATGCGGAGCAG 2437
Qy      263 HisThrAsnGlyPheValSerSerTyrlleHisIleLysAsp---AlaGlnValLysThr 281
Db      2436 ACCCGTCGCGGACCGAGCTCCAGGAAGGTCTCGCGCGCGCATTCCTCCAGGTGCTG 2377
Qy      282 GtAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 301
Db      2376 GCGTCGAGGAAGGCTACGGGCGCTGCGG-----ACCTGTCACACCGATCTCGGATCG 2323

RESULT 16
US-09-861-289-1/c
/ Sequence 1, Application US/09861289
/ Patent No. US20020110897A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/861,289
/ PRIOR FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: 09/105,537
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 1
/ LENGTH: 15872
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Alignment Scores:
Pred. No.: 0.00547 Length: 15872
Score: 130.50 Matches: 81
Percent Similarity: 38.75% Conservative: 43
Best Local Similarity: 25.31% Mismatches: 115
Query Match: 7.81% Indels: 81
Gaps: 18

US-10-018-706-2 (1-322) x US-09-861-289-1 (1-15872)
Qy      34 AlaSerLysProThrTyraSerThrSerGlySerHisArgThrSerGlySer 53
Db      3195 GCGTCGCGCGCC-----TCGGCAGCAGCGCCGACAGGTCTCGGGGAG 3151
Qy      54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
Db      3150 GTGACCGCGCGGGGTAGCGGAGCCGATGCCATGATTCGCGATGCGGTGCGGCGGTG 3091
Qy      64 ThrAsp-----SerGlnGlyValProAsnArgTyrglnValLysGlnIleAspThrVal 81

```

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rnpb

Page 14

```
Db 3090 TCCGCGGAGGTCGCGGCGGATTCCTCCGCGGATTCGCGGCTGCGCGGCGGTCG 3031
QY 82 SerIys-----IleAlaGlnArgIleuAsnTrpArgIle----- 94
Db 3030 AGCAGGTCGCCAGAGTCGCGGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 94
QY 95 IleGlyHisIleAsnMetLeuAsnSerIleTrpIleTrpIleGlyValIleTrpLeuThr 2971
Db 2970 CTGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2971
QY 115 LeuTrpSerGlyAsp-----LeuIleValIleArgSer 126
Db 2925 -----TCGACGAGATCGAGGAGTCGAGGAGCCGCGGCTCTGAGAGTGGCGGAGTTCG 2872
QY 127 IleserSerGly-----ValAsnThrAlaHisThrProSerPro 139
Db 2871 AGCGGAGTCGCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2812
QY 140 ValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValAlaGlnIleProThr 2812
Db 2811 GCTCGTGGCGGAGCGCGGCGGATCACCG-----CCACCGGACCGCGCGCGCTCG 2764
QY 160 ProProValValValValValValValValValValValValValValValValValVal 2764
Db 2763 CCC-----TCGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 2722
QY 180 ProValAlaProProValThrGlnAlaProPheAlaThrGlySer----- 194
Db 2721 CCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 194
QY 195 -----SerGlyValMetGlnPheArgIlePro-----ValGlyAlaThrAsnProValIle 2662
Db 2661 CCGTGAACAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 2602
QY 212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValIleThr 225
Db 2601 CCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 2542
QY 226 Ser-----AsnGlyMetIleTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db 2541 AGTGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2497
QY 244 AlaGlyThrValIleGlnAlaAsp-----HisAsnMetAspGlyAlaSerIleValIleGln 262
Db 2496 GCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 2437
QY 263 HisThrAsnGlyPheValSerSerIleHisIleIleAsp-----AlaGlnValIleThr 281
Db 2436 ACCCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2377
QY 282 GlyAspThrValArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyAla 301
Db 2376 GCGTGGAGAGAGCGTACCGGCGGCGGCGG-----ACCTGCTCACCGAGTACTCGGAGTCG 2323
RESULT 17
US-10-087-667-4
Sequence 4, Application US/10087667
Publication No. US20020194629A1
GENERAL INFORMATION:
APPLICANT: Bramley, John A.
APPLICANT: Plaut, Karen I.
TITLE OF INVENTION: TREATMENT OF STAPHYLOCOCCUS INFECTIONS
FILE REFERENCE: Masclis
CURRENT FILING DATE: 2002-02-28
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 1520
TYPE: DNA
ORGANISM: Staphylococcus simulans
```

US-10-087-667-4

Alignment Scores:

```
Score: 0.00038
Percent Similarity: 128.50
Best Local Similarity: 33.98%
Query Match: 22.01%
DB: 7.69%
Gaps: 15
```

US-10-018-706-2 (1-322) x US-10-087-667-4 (1-1520)

```
QY 10 GlnAsnGlnIleProIleLysArgLeuGlyLeuIlePheGlyValIleThrThrCysIle 29
Db 313 CAGCATATGAGAGATTCACAGCGGCGGACTG-----GGCTGGCGGCTGCTGCTGCGG 366
QY 30 LeuAlaGlyCysAlaSerIleProThrTyraSerThrSerIleSerIleSerIleArg 49
Db 367 CTGGCG-----ACGATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 399
QY 50 ThrSerGlySerGlyIleValAlaIleGlySerGlnValIleThrAspSerGlnIleVal 69
Db 400 GCCACGGCTCAG-----CGCGGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 444
QY 70 ProAsnArgIleValIleValIleValIleValIleValIleValIleValIleValIle 444
Db 445 -----TTGCACTTCGACATCGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
QY 90 LeuAsnTrpArgIleGlyHisIleAsnMetLeuAsnSerSerIleTrpIleThr 109
Db 496 AAGCACTCGGAGAGATCTGCG-----TACAGCGGAGATCGGCGGCGGCGGCGGCGG 519
QY 110 GlyIleTrpLeuThrIleThrIleThrIleThrIleThrIleThrIleThrIleThrIle 129
Db 520 -----TGGCGCGC-----TACAGCGGAGATCGGCGGCGGCGGCGGCGGCGG 546
QY 130 GlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProVal 149
Db 547 AGTGTATCGCGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606
QY 150 GlnGlnHisProAlaValGlnIleProThrProProValValValIleValIlePro 168
Db 607 GTCGTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 657
QY 169 -----ThrProThrProProValIleGln----- 176
Db 658 AGTGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 717
QY 176 ----- 176
Db 718 GGTAGCGTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 777
QY 177 -----GlnProAlaProValAlaProProValIleThr 186
Db 778 GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 837
QY 187 Gln-----AlaPro-----Phe 190
Db 838 CGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 897
QY 191 AlaThrGlySerSerGlyValMetGlnPheArgIleProValGlyValAlaHisProVal 210
Db 898 GCGGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 957
QY 211 ValArgArgPheGlyThrAlaThrValAlaGlySer----- 222
Db 958 -----GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 222
QY 223 ThrValThrSerAsnGlyMetIleTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSer 242
Db 1009 GACATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1068
QY 243 AsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGln 262
```

```
Db 1069 GCCCGCGCTGCTCAACGCG-----CACTTCGTGCTTCGCGGAGATCGTG----- 1116
Qy 263 HlthraanglypnevalserSerTyrllehllelysaaplaaglnvallyserGly 282
Db 1117 CACACCGCGCTGCTGACGACCACTTACCACTGATGAACCTTACCACTGATGAAC 1176
Qy 283 AsphrValarglyThrGlylnArglylealaserMetlysaanglnProserGlyAla 301
Db 1177 GCCACGTTGTCGATGACACCGCCATCGCCCAACCGCCCAACCCAGGCGAGCGC 1233

RESULT 18
US-09-468-147-164
; Sequence 164, Application US/09468147A
; Publication No. US20030049601A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Schlander, George G.
; APPLICANT: Erker, James C.
; APPLICANT: Desai, Suresh M.
; APPLICANT: Dawson, George J.
; APPLICANT: Mushawar, I. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; FILE REFERENCE: 6232.US.P1
; CURRENT APPLICATION NUMBER: US/09/468,147A
; EARLIER FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: US 09/173,141
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: US 60/061,199
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: PasteSeq for Windows Version 3.0
; SEQ ID NO 164
; LENGTH: 7277
; TYPE: DNA
; ORGANISM: Hepatitis E Virus
; FEATURE:
; OTHER INFORMATION: us2full1
US-09-468-147-164

Alignment Scores:
Pred. No.: 0.00438 Length: 7277
Score: 127.00 Matches: 80
Percent Similarity: 33.14% Conservative: 37
Best Local Similarity: 22.66% Mismatches: 120
Query Match: 7.60% Indels: 116
DB: 9 Gaps: 19

US-10-018-706-2 (1-322) x US-09-468-147-164 (1-7277)
Qy 27 ThrCyaleleuAlaGlyCyA-----AlaserlyspProThrTyraanserThrs 43
Db 1755 ACGAGCGGTGTGAATGAGCGCCATCTTGAAGCGAATGCGCCCTGAGAGTATGTTCTGCA 1814
Qy 44 GlyserGlyserHiaArgThrseryserGlyGlyleuAlaIle----- 58
Db 1815 TTGAAGCCTCTCGCGAGCTATGAGGCGCGGCGTGCACAGAGCTTACTTATGAGTCAACC 1874
Qy 59 -----GlyserGlnValIleThrsaserglnGlyAlaProasArgTyrglnVallys 76
Db 1875 CCGTGGCGGTTCGAGGTAAAGTTTCACTTATGATGCTGAGTGCACGCAATTCCTCC 1934
Qy 77 GlnGlyasprThrsValser----- 82
Db 1935 CCYGGTGGCGCCCTTAGCGCGCGCGGAGAGTGAGCGCCTTCTGACAGTCTTTAT 1994
Qy 83 -----LysIleAlaGlnArgTyrglyleuAan-----TrrArg----- 93
Db 1995 AGATCAATAGTGTACCCAGCGGCACTTCGTCGACAGGCGGACATAGGCTACCTCTGAG 2054
Qy 94 -----GlnleGlyHleIle-----AsnAan 100
```

```
Db 2055 GGGCTGCTGGGTACTTTCCTCCCACTTCTCCCTGGGAGATTTGGAGTCTGTACCCC 2114
Qy 101 LeuanserSerTyrlleTyrlleThrglyGlnTrpLeuThrleu---TrrSerGlyAsp 119
Db 2115 TTTTGGGTGAGGGGACTTTGTATACCGGACCTGACCACTTCTGTTTCTTCTAGTAT 2174
Qy 120 LeuValarglylnArglysaerIleSerSerGlyValasnthralaHlthrsProserPro 139
Db 2175 TTTTCCCCCTTACGAGGCGGCGCTTCCTTCGCTTCGCTCCGCGCGGAGTTGCCCTACCT 2234
Qy 140 ValAlaValGlnserSerArgProVal----- 149
Db 2235 ACT-----CCACTGTATGATATCTGGGTATTACACCGCCCTCA 2276
Qy 150 -----GlnGlnHls-----ProAlaValGlnlysPro-----Thr 159
Db 2277 GAGGAATCTCATGTGTATGAGGCGCATCTGTACCTCTCTCTCTGAGCCCTGCTGATGAC 2336
Qy 160 ProProValValValVallylsProThrProThrProProValValGlnGlnProAla 179
Db 2337 AGCCCTATTGTGCTTACC-----CCCCCCCCCCCCCTCTCCGTCGTAACCGGCA 2390
Qy 180 ProValAlaProProValThrglnAlaProPheAlaThrglySerSerGlyValMetGln 199
Db 2391 ACATCCCCGCTCCCGCCACTCGCCCT-----CTCCTT 2423
Qy 200 PheArgTyrlleProValGlyAlaThraAsnProValValArgArgPheGlyThraAlaThraVal 219
Db 2424 TACACCTACCCCGACGCGCGC-----AACGTGTAT 2453
Qy 220 AlaGlySerThrsValThrsAsnGlyMetTrpPheSerSerGlyArgAspGlyAspLeuIle 239
Db 2454 GCGGGGTCACTTGTGTGATGACAGTGTGATGG-----TTTGTGTC 2492
Qy 240 AsnAlaserAsnAlaGlyThrsValIleGlnAlaAspHisanMetAspGlyAlaSerIle 259
Db 2493 AATGCTCAAAACCTGTGC-----CATCGCCCCGGGGGTGC----- 2528
Qy 260 ValIleGlnHlsThraanglypnevalserSerTyrllehllelysaaplaaglnVal 279
Db 2529 ---CTGTGCAT-----GCTTTTATCAAGCTTTCAGAGAGCGTTCTTAC 2570
Qy 280 LysThrglyasprThrsValargThrglyGlnArglylealaserMetlysaanglnPro--- 298
Db 2571 TCGACTGAATTCATCATGCGCGGAGGCGCTTGCAGACATACCTTTAACCCCGCCCTATT 2630
Qy 299 SerGlyAlaAlaLeupheGlnPheArglyleaserArgAan 311
Db 2631 ATCCATGCAAGTGTCCCGCACTATGAGTTGAGCAAAAC 2669

RESULT 19
US-09-468-147-165
; Sequence 165, Application US/09468147A
; Publication No. US20030049601A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Schlander, George G.
; APPLICANT: Erker, James C.
; APPLICANT: Desai, Suresh M.
; APPLICANT: Dawson, George J.
; APPLICANT: Mushawar, I. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; FILE REFERENCE: 6232.US.P1
; CURRENT APPLICATION NUMBER: US/09/468,147A
; EARLIER FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: US 09/173,141
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: US 60/061,199
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: PasteSeq for Windows Version 3.0
; SEQ ID NO 165
```



```

1 LENGTH: 7277
2 TYPE: DNA
3 ORGANISM: Hepatitis B Virus
4 FEATURE:
5 OTHER INFORMATION: us2full1
6 FEATURE:
7 NAME/KEY: CDS
8 LOCATION: (36)...(5159)
9 OTHER INFORMATION: orf1
10 FEATURE:
11 NAME/KEY: CDS
12 LOCATION: (5197)..(7176)
13 OTHER INFORMATION: orf2
14 OTHER INFORMATION: orf3
15 OTHER INFORMATION: Xaa = unknown or Other at position 5159-5527
16 OTHER INFORMATION: Xaa = unknown or Other at position 322
17 OTHER INFORMATION: Xaa = unknown or Other at position 331
18 OTHER INFORMATION: Xaa = unknown or Other at position 445
19 OTHER INFORMATION: Xaa = unknown or Other at position 448
20 OTHER INFORMATION: Xaa = unknown or Other at position 634
21 OTHER INFORMATION: Xaa = unknown or Other at position 646
22 OTHER INFORMATION: Xaa = unknown or Other at position 811
23 OTHER INFORMATION: Xaa = unknown or Other at position 1553
24 OTHER INFORMATION: Xaa = unknown or Other at position 1578
25 OTHER INFORMATION: Xaa = unknown or Other at position 1597
26 OTHER INFORMATION: Xaa = unknown or Other at position 1792
27 OTHER INFORMATION: Xaa = unknown or Other at position 1938
28 OTHER INFORMATION: Xaa = unknown or Other at position 2155
29 OS-09-468-147-165

```


TITLE OF INVENTION: HEPATITIS E VIRUS
 FILE REFERENCE: 6232 US P1
 CURRENT APPLICATION NUMBER: US/09/468,147A
 CURRENT FILING DATE: 1999-12-21
 EARLIER APPLICATION NUMBER: US 09/173,141
 EARLIER FILING DATE: 1998-10-15
 EARLIER APPLICATION NUMBER: US 60/061,199
 EARLIER FILING DATE: 1997-10-15
 NUMBER OF SEQ ID NOS: 256
 SOFTWARE: Fastsq for Windows Version 3.0
 SEQ ID NO 141
 LENGTH: 903
 TYPE: DNA
 ORGANISM: Hepatitis E virus
 FEATURES:
 OTHER INFORMATION: us2-851
 US-09-468-147-141

Alignment Scores:
 Pred. No.: 0.000388 Length: 903
 Score: 125.00 Matches: 61
 Percent Similarity: 35.95% Conservative: 26
 Best Local Similarity: 25.21% Mismatches: 81
 Query Match: 7.49% Indels: 74
 Gaps: 14

US-10-018-706-2 (1-322) x US-09-468-147-141 (1-903)

96 GlyHisle-----AaAnleuAnSerSerrTyrrHleTyrrHlyGln 111
 205 GGGCAATTTTGGAGAGTCTGTAACCCCTTTGGCGGAGGAGGACTGTATACCCGAC 264
 112 TrrleuHleu---TrrSerGlyAepLeuValArgGluArgSerIleSerSergly 130
 265 TGGTCAACCTCTGCTTTTCTAGTATTTCTCCCTAGGCGCGCTCTCTGCTTG 324
 131 ValAsnThralHisThrProSerProValAlaValGlnSerSerrArgProVal 149
 325 GCTGCGCGCGCGGAGTTCCTTACCTTACT-----CCACCTGTACT 366
 150 -----GlnGlnHis-----Pro 153
 367 GATATCTGGGTGTACACCGCCCTCAGAGATCTCATGTTGATGCGCATCTGACCC 426
 154 AlaValGlnlyPro-----ThrProValAlaValAlaValysProThrPro 170
 427 TCTGTTCTGAGCGCTGCTGATGACACGCGCTATGTGCTTAC-----CCCCCCCC 480
 171 ThrProValAlaGlnGlnProAlaProValAlaProValAlaProHe 190
 481 CCCCCTCTCCCTGCGTACGCGACATCCCGCTCCCGCAGCTGCGCGT 534
 191 AlaThrGlySerSerglyValMetGlnPheArgTyrrProValGlyAlaThrAnProVal 210
 535 -----CTCCTTACACTTACCCGACGCGCC-----561
 211 ValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAnGlyMetTrp 230
 562 -----AAGTGATGCGGGGCTGTGTGAGCAACAGCTGATGAG 603
 231 PheSerGlyArgArgPheGlyAepLeuIleAnAlaSerAnAlaGlyThrValIleGlnAla 250
 604 -----TTAGTCATAGCTTCAACCTCGC-----627
 251 AspHisAsnMetAepGlyAlaSerIleValIleGlnHisThrAnGlyPheValSerSer 270
 628 ---CATCGCCCGCGGGGTGAC-----CTCTGCAAT-----GCTTTT 660
 271 TyrrleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArg 290
 661 TATTCACGTTTCCCAAGAGCTTCTACTGACTGATTCATCATGCGCGAGGCGCTTGA 720
 291 IleAlaSerMetLysAnGlnPro---SerglyAlaIleLeuPheGlnPheArgIleSer 309

Db 721 GCATACACTTAAACCCGCGCTTATTCATCATGAGTGGCTCCCGACTATAGGTGAG 780
 310 ArgAn 311
 781 CAANAAC 786
 RESULT 23
 US-10-063-547-99
 Sequence 99, Application US/10063547
 Publication No. US20020182638A1
 GENERAL INFORMATION:
 APPLICANT: Eaton, Dan L.
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Gerltzen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Grimaldi, Christopher J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Marabe, Colin K.
 APPLICANT: Wood, William I.
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3230P1C1
 CURRENT APPLICATION NUMBER: US/10/063,547
 PRIOR APPLICATION removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 170
 SEQ ID NO 99
 LENGTH: 2436
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-10-063-547-99
 Alignment Scores:
 Pred. No.: 0.00289 Length: 2436
 Score: 122.50 Matches: 78
 Percent Similarity: 39.04% Conservative: 52
 Best Local Similarity: 23.42% Mismatches: 145
 Query Match: 7.34% Indels: 60
 Gaps: 12
 US-10-018-706-2 (1-322) x US-10-063-547-99 (1-2436)
 6 AalIleAnSerGlnAnGlnlyProIleLysArgLeuGlyLeuIlePheGlyValIle 25
 686 GCGACCAACTGTAGTCCAGACAGTCCAGTACGCGCCAGCACTGCCAACACTGAG 745
 26 ThrThrCysIleLeuAlaGlyCysAlaSerlyProThrTyrrAnSerThrSerglySer 45
 746 TCTACACACTCTCCAGTGGGCG-----AGACAGGACCAACTCT 787
 46 GlySerHisArgThrSerglySerglyGlyLeuAlaIleGlySerGlnValIleThrAsp 65
 788 GACTTCAGCACACACTCCTCAGTGGGCTTAGCAACCACTCTGATGTCAGACAAACC 847
 66 SerGlnGlyValProAnArgTyrrGln-----ValLysGlnGlyAsp 79
 848 TCCAGTGGGGCGAG---CACACCAACCACTGAGTCCAGACAGAGTCCAGTGGGCGAG 906
 80 ThrValSerlySrlleAlaGlnArgTyrrGlyLeuAnThrPheGlnIleGlnHis-----97
 907 CACTGCCACCAACTCTAGTCCAGACCACTCCTCAGTGGGCGAGCAAGCCCAACTC 966
 98 -----IleAnAnleu---AnSerSerrTyrrHleTyrrHlyGlnThrPheLeu 115
 967 TGAATCCAGAACCGACTCCATATGGGCTGGCAAGCCCAACTCTGATCCAGACGAC 1026
 115 utrSerglyAepLeuLysValArg-----GluArgSerIleSerSerglyValAs 132
 1027 CTCAGTGGGGCGAGCAGACAGCAACCACTGACTCCAGACAGAGTCTCAGTGGGCGAG 1086
 132 nThrAla-HisThrProSerProValAlaValGlnSerSerrArgProProVal-----G 150

```

Db      1087 CACTGCCACCACTGTGAGTCCA-----GCACGACCTCCAGTGGGGCCAG 1131
Qy      150 IngInHisProAlaValGlnInysProThrProProValValValValysIserProThrP 170
Db      1132 CACAGCCACCACTGTGAGTCCAGACGACGACGAGTGGGGCTGAGCCAGCCACCACTC 1191
Qy      170 roThrProProValValGlnInPro--AlaProValAlaProProVal-----T 186
Db      1192 TGACTCCA-----GCACACTCTCCAGTGGGGCCGGCAGCAGCCACTGTGAGTCCAG 1245
Qy      186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db      1246 CACAGTGTCCAGTGGATCAGACAGTCCACCAATCTGAGTCCAGACACCTCC-AGTG 1304
Qy      206 laThrAsnProValValArgArgPheGlyThrAlaAlaGlySerThrValThrs 226
Db      1305 GGGCCAC------ACAGCCACCACTCTGAGTCCAGTCCAGTCCAGTCC 1343
Qy      226 erAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db      1344 CCAGTGG------GCACACACAGCCA 1364
Qy      246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
Db      1365 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGACAGTCCACCACTCTGAGTCCA 1424
Qy      264 -----ThraSnglyPheValSerSerTyrIleHisIleLysAspAlaGlnValysThrG 282
Db      1425 GCACACCTCCAGTGGGGTCCAGCAGCAGCCACCACTGTGAGTCCAGCACAACCTCCAGT 1484
Qy      282 lyaAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaA 302
Db      1485 GGGGTAGACAGCAGCAGCACTCTGACTCTGACTCCAGCAGCACTCCAGTGGCCAGCAGCCA 1544
Qy      302 laLeuPheGlnPheArgIleSerArgAsnGlyVal 313
Db      1545 CCAACTGTGAGTCCAGCAGTGTCCAGTGGGATC 1579

RESULT 24
US-10-174-590-309
: Sequence 309, Application US/10174590
: Publication No. US20030008352A1
: GENERAL INFORMATION:
: APPLICANT: Baker Kevin P.
: APPLICANT: Chen, Dian
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3430R1C42
: CURRENT APPLICATION NUMBER: US/10/174,590
: PRIOR FILING DATE: 2002-06-18
: Prior application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 309
: LENGTH: 2436
: TYPE: DNA
: ORGANISM: Homo Sapien
US-10-174-590-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservativeness: 52
Best Local Similarity: 23.42% Mismatches: 145

```

```

Query Match: 7.34% Indels: 60
DB: 9 Gaps: 12
US-10-018-706-2 (1-322) x US-10-174-590-309 (1-2436)
Qy      6 AlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeuIlePheGlyValIle 25
Db      686 GCCACCAACTCTGAGTCCAGACAGTGTCCAGTGGGCGAGCAGTCCACCAACTGTAG 745
Qy      26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrTrpAsnSerThrSerGlySer 45
Db      746 TCTAGCAGCTCTCCAGTGGGCC-----AGCAGGCCACCACTCT 787
Qy      46 GlySerHisArgThrSerGlySerGlyLysLeuAlaIleGlySerGlnValIleThrAsp 65
Db      788 GACTCCAGCAGCAGCAGTCCAGTGGGCTGAGCAGCAGCAGCAGCAGTCCAGTCCAGCAGCAG 847
Qy      66 SerGlnGlyValProAsnArgTyrGln-----VallyGlnGlyAsp 79
Db      848 TCCAGTGGGGCCAG-CACAGCCACCAACTCTGAGTCCAGCAGCAGTGTCCAGTGGGCCAG 906
Qy      80 ThrValSerLysIleAlaGlnArgTyrGlyLysAsnTrpArgGluIleGlyHis----- 97
Db      907 CACTGCCACCAACTCTGAGTCCAGCAGCAGCAGTCCAGTGGGGCCAGCAGCAGCAGCAGTCC 966
Qy      98 -----IleAsnAsnLeu-AsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThrIle 115
Db      967 TGAGTCCAGAAAGCACTCTCAATGGGGCTGGCAGCAGCCACCACTGTGAGTCCAGCAGCAG 1026
Qy      115 utrPserGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
Db      1027 CTCAGTGGGGCCAGCAGCAGCAGCAGCAGTCCAGTCCAGCAGTGTCCAGTGGGGCCAG 1086
Qy      132 ntThrAla-HisTrpProSerProValAlaValGlnSerArgProProVal-----G 150
Db      1087 CACTGCCACCAACTCTGAGTCCA-----GCACGACCTCCAGTGGGGCCAG 1131
Qy      150 IngInHisProAlaValGlnLysProThrProProValValValValysIserProThrP 170
Db      1132 CACAGCCACCAACTGTGAGTCCAGCAGCAGCAGTCCAGTGGGGCTGAGCCAGCCACCACTC 1191
Qy      170 roThrProProValValGlnInPro--AlaProValAlaProProVal-----T 186
Db      1192 TGACTCCA-----GCACACTCTCCAGTGGGGCCGGCAGCAGCCACTGTGAGTCCAG 1245
Qy      186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db      1246 CACAGTGTCCAGTGGATCAGACAGTCCACCAATCTGAGTCCAGCAGCAGCAGTCC-AGTG 1304
Qy      206 laThrAsnProValValArgArgPheGlyThrAlaAlaGlySerThrValThrs 226
Db      1305 GGGCCAC------ACAGCCACCAACTCTGAGTCCAGTCCAGTCCAGTCCAGTCC 1343
Qy      226 erAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db      1344 CCAGTGG-----GCACACACAGCCA 1364
Qy      246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
Db      1365 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGCAGTCCACCACTCTGAGTCCA 1424
Qy      264 -----ThraSnglyPheValSerSerTyrIleHisIleLysAspAlaGlnValysThrG 282
Db      1425 GCACACCTCCAGTGGGGTCCAGCAGCAGCCACCACTGTGAGTCCAGCACAACCTCCAGT 1484
Qy      282 lyaAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaA 302
Db      1485 GGGGTAGACAGCAGCAGCACTCTGACTCTGACTCCAGCAGCACTCCAGTGGCCAGCAGCCA 1544
Qy      302 laLeuPheGlnPheArgIleSerArgAsnGlyVal 313
Db      1545 CCAACTGTGAGTCCAGCAGTGTCCAGTGGGATC 1579

```

RESULT 25
 US-10-176-758-309
 ; Sequence 309, Application US/10176758
 ; Publication No. US20030008353A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P1430P1C104
 ; CURRENT APPLICATION NUMBER: US/10/176,758
 ; CURRENT FILING DATE: 2002-06-21
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 309
 ; LENGTH: 2436
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-176-758-309

Alignment Scores:
 Pred. No.: 0.00289 Length: 2436
 Score: 122.50 Matches: 78
 Percent Similarity: 39.04% Conservative: 52
 Best Local Similarity: 23.42% Mismatches: 145
 Query Match: 7.34% Indels: 60
 DB: Gaps: 12

US-10-018-706-2 (1-322) x US-10-176-758-309 (1-2436)

```

QY 6 AAlaIleasnSerGlaSndInlySProIlelySArGleuGlyLeuIlePheGlyValIle 25
DB 686 GCCACCACTCTGAGTCCAGCAGAGTCCAGTCCAGTGGGCGACAGCTGCCACCACTCTGAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerIlySProThrIlyrAsnSerThSergIlySer 45
DB 746 TCTAGCACCTCTCCAGTGGGCC-----AGCACGCCACCACTCT 787
QY 46 GlySerHisArgThSergIlySergIlyLeuAlaIleGlySergIlyValIleThrAsp 65
DB 788 GACTCCAGCAGCAACCTCTGAGTGGGCTAGCAGCAGCAGCACTCTGATCCAGCAGCAACC 847
QY 66 SerGlnGlyValProhennArgIlyrGln-----VallyGlnGlyAsp 79
DB 848 TCCAGTGGGGCCAG-CACAGCCACCACTCTGATGTCGACAGCAGTGTCCAGAGGGCCAG 906
QY 80 ThrValSerIlyleAlaGlnArgIlyrGlyLeuAsnIlyrParGlnIleGlyHis----- 97
DB 907 CACTGCCACCACTCTGAGTCCAGCAGCACTCTGAGTGGGCGACAGTGTCCAGCAACTC 966
QY 98 -----IleAsnAsnLeu-AsnSerSerIlyrThrIleIlyrThrGlyGlnIlyrPleuThrIle 115
DB 967 TGAGTCCAGACGACCTCCATGAGGGCTGGCAGCAGCAGCAACTCTGATCCAGCAGCAGC 1026
QY 115 UTrpSerGlyAspIleLeuValArg-----GluArgSerIleSerSergIlyValAs 132
DB 1027 CTCACAGTGGGGCCAGCAGCAGCAGCACTCTGATCCAGCAGCAGTGTCCAGTGGGGCCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerProProVal-----G 150
DB 1087 CACTGCCACCACTCTGAGTCCA-----GCAGACCTCCAGTGGGGCCAG 1131
QY 150 InGlnHisProAlaValAlaGlnlySProThrProProValValValIlySlySProThrP 170
  
```

DB 1132 CACAGCCACCAACTCTGAGTCCAGCAGCACTCTCCAGTGGGCTAGCAGCAGCAGCAACTC 1191
 QY 170 rOfThProProValValInGlnPro---AlaProValAlaProProVal-----T 186
 DB 1192 TGACTCCA-----GCACAACTCTCCAGTGGGGCGGCGACAGCCACCACTCTGATCCAG 1245
 QY 186 hGlnAlaProPheAlaThrGlySerSergIlyValMetGlnPheArgIlyrProValAla 206
 DB 1246 CACAGTGTCCAGTGGGATCCAGCAGCAGTCCCAATTTCTGAGTCCAGCAGCAGCTCC-AGTG 1304
 QY 206 lArhAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValIlys 226
 DB 1305 GGGCCAAAC-----ACAGCCACCACTCTGAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 1343
 QY 226 eRAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
 DB 1344 CCAGTGGG-----GCCAACACAGCCA 1364
 QY 246 hVAlIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
 DB 1365 CCACTCTGAGTCCAGCAGCAGTGTCCAGTGGGGCGAGCAGTCCACCACTCTGATCCAGTCCA 1424
 QY 264 -----ThrAsnGlyPheValSerSeryrIleHisIleLeuAspAlaGlnValIlyrThg 282
 DB 1425 GCACAACTCTCCAGTGGGCTCAGCAGCAGCAGCACTCTGATCCAGCAGCAACTCTCAGTG 1484
 QY 282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetIlyAsnGlnProSergIlyAla 302
 DB 1485 GGAGTTCAGCAGCAGCCACCACTCTGATCCAGCAGCACTCTCAGTGGAGCCAGCAGCAGCA 1544
 QY 302 lAlaPheGlnPheArgIleSerArgAsnGlyVal 313
 DB 1545 CCAACTCTGAGTCTAGCAGCAGTGTCCAGTGGAGTCC 1579

RESULT 26
 US-10-063-616-99
 ; Sequence 99, Application US/10063616
 ; Publication No. US20030013855A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P1230P1C1
 ; CURRENT APPLICATION NUMBER: US/10/063,616
 ; CURRENT FILING DATE: 2002-05-03
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 170
 ; SEQ ID NO 99
 ; LENGTH: 2436
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-063-616-99

Alignment Scores:
 Pred. No.: 0.00289 Length: 2436
 Score: 122.50 Matches: 78
 Percent Similarity: 39.04% Conservative: 52
 Best Local Similarity: 23.42% Mismatches: 145
 Query Match: 7.34% Indels: 60
 DB: Gaps: 12

US-10-018-706-2 (1-322) x US-10-063-616-99 (1-2436)

```

QY 6 AAlaIleasnSerGlaSndInlySProIlelySArGleuGlyLeuIlePheGlyValIle 25
  
```

```

Db      686 GCCACCACTGTGAGTCAGACAGAGTCGATGAGGCGACAGCTGCCACCAACTGTAG 745
Qy      26 ThrThrCysIleLeuAlaGlyCysAlaSerIleProThrTyraSerThrSerylSer 45
Db      746 TCTACACACTCTCCAGTGGGGCC-----AGCAGCCACCACTCT 787
Qy      46 GlySerHisArgThrSergIleSergIleValIleGlySergIleValIleThrAsp 65
Db      788 GACTCCAGACACACTCCAGTGGGGCTGACAGCAGCACTGAGTCCAGACCAACC 847
Qy      66 SerGlnGlyValProAsnArgTyrln-----VallyGlnGlyAsp 79
Db      848 TCCAGTGGGGCCAG-CACAGCCACCACTGAGTCCAGCAGCAGTGTCCAGTGGGGCCAG 906
Qy      80 ThrValSerIleValIleAlaGlnArgTyrlnLeuAsnTrpArgGluIleGlyHis----- 97
Db      907 CACTGCCCACTGTGAGTCAGCAGCACTCCAGTGGGGCCAGCAGCAGCAGCAGCAGCTC 966
Qy      98 -----IleAsnAsnLeu-AsnSerSerTyrlleTyrlleTyrlleGlnTrpLeuThrle 115
Db      967 TGAAGTCAGAAAGCACTCCAAATGGGGCTGGGACAGCCAGCACTGTGAGTCCAGCAGCAG 1026
Qy      115 UTPrSergIleAspLeuIleValArg-----GluArgSerIleSergIleValAs 132
Db      1027 CTCAGTGGGGCCAGCAGCAGCAGCACTGAGTCCAGTCCAGTGTCCAGTGGGGCCAG 1086
Qy      132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db      1087 CACTGCCCACTGTGAGTCA-----GCAGAGCTCCAGTGGGGCCAG 1131
Qy      150 lnglnHisProAlaValGlnIleProThrProProValAlaValIleValIleProThrP 170
Db      1132 CACAGCCCACTGTGAGTCCAGCAGCAGCACTCCAGTGGGGCTGACAGCAGCAGCACTC 1191
Qy      170 roThrProProValValGlnIlePro--AlaProValAlaProProVal-----T 186
Db      1192 TGACTCCA-----GCACAACTCCAGTGGGGCCAGCAGCAGCACTGTGAGTCCAG 1245
Qy      186 hrGlnAlaProPheAlaThrGlySerSergIleValMetGlnPheArgTyrlleProValGly 206
Db      1246 CACAGTGTCCAGTGGGATGACAGCAGTCAATCTGAGTCCAGCAGCACTCCAGTCC-AGTG 1304
Qy      206 lArAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValIleThr 226
Db      1305 GGGGCCAAC-----ACAGCCACCACTGTGAGTCCAGTCCAGTCCAGT 1343
Qy      226 eraGlnGlyMetTrpPheSergIleArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db      1344 CCAGTGGG-----GCCAACACAGCCA 1364
Qy      246 hrValIleGlnAlaAspHisAsnMetArgIleValIleValIleGlnHis----- 263
Db      1365 CCAACTGTGAGTCCAGCAGAGTGTCCAGTGGGGCCAGCAGTCCAGCACTGTGAGTCCA 1424
Qy      264 -----ThrAsnGlyPheValIleSerSerTyrlleHisIleGlyAspAlaGlnValIle 282
Db      1425 GCACAACTCCAGTGGGGCTGACAGCAGCAGCACTGAGTCCAGCAGCACTCCAGT 1484
Qy      282 lyaSerThrValArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSergIleVal 302
Db      1485 GGGGTACAGCAGCCCACTGTGAGTCCAGCAGCACTCCAGTGGGGCCAGCAGCAGCA 1544
Qy      302 lAlaLeuPheGlnPheArgIleSerArgAsnGlyVal 313
Db      1545 CCAACTGTGAGTCCAGCAGAGTGTCCAGTGGGATC 1579

```

```

; APPLICANT: Desnoyers, Inc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; PRIORITY FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 309
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-175-737-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: Gaps: 12

US-10-018-706-2 (1-322) x US-10-175-737-309 (1-2436)
Qy      6 AlaIleAsnSerGlnAsnGlnIleProIleValArgLeuGlyLeuIlePheGlyValIle 25
Db      686 GCCACCACTGTGAGTCCAGCAGAGTGTCCAGTGGGGCCAGCAGCAGTCCAGCACTGTAG 745
Qy      26 ThrThrCysIleLeuAlaGlyCysAlaSerIleProThrTyraSerThrSerylSer 45
Db      746 TCTACACACTCTCCAGTGGGGCC-----AGCAGCCACCACTCT 787
Qy      46 GlySerHisArgThrSergIleSergIleValIleGlySergIleValIleThrAsp 65
Db      788 GACTCCAGACACACTCCAGTGGGGCTGACAGCAGCAGCACTGAGTCCAGCAGCAACC 847
Qy      66 SerGlnGlyValProAsnArgTyrln-----VallyGlnGlyAsp 79
Db      848 TCCAGTGGGGCCAG-CACAGCCACCACTGAGTCCAGCAGCAGTGTCCAGTGGGGCCAG 906
Qy      80 ThrValSerIleValIleAlaGlnArgTyrlnLeuAsnTrpArgGluIleGlyHis----- 97
Db      907 CACTGCCCACTGTGAGTCCAGCAGCACTCCAGTGGGGCCAGCAGCAGCAGCAGCAGCTC 966
Qy      98 -----IleAsnAsnLeu-AsnSerSerTyrlleTyrlleTyrlleGlnTrpLeuThrle 115
Db      967 TGAAGTCAGAAAGCACTCCAAATGGGGCTGGGACAGCCAGCACTGTGAGTCCAGCAGCAG 1026
Qy      115 UTPrSergIleAspLeuIleValArg-----GluArgSerIleSergIleValAs 132
Db      1027 CTCAGTGGGGCCAGCAGCAGCAGCACTGAGTCCAGCAGCAGTGTCCAGTGGGGCCAG 1086
Qy      132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db      1087 CACTGCCCACTGTGAGTCA-----GCAGAGCTCCAGTGGGGCCAG 1131
Qy      150 lnglnHisProAlaValGlnIleProThrProProValAlaValIleValIleProThrP 170
Db      1132 CACAGCCCACTGTGAGTCCAGCAGCAGCACTCCAGTGGGGCTGACAGCAGCAGCACTC 1191
Qy      170 roThrProProValValGlnIlePro--AlaProValAlaProProVal-----T 186
Db      1192 TGACTCCA-----GCACAACTCCAGTGGGGCCAGCAGCAGCACTGTGAGTCCAG 1245
Qy      186 hrGlnAlaProPheAlaThrGlySerSergIleValMetGlnPheArgTyrlleProValGly 206

```

```

Db      1246 CACAGTGTCTGAGTGGATGACGACAGTCCAAATTCATGTCAGCACACCTCTCC-AGTG 1304
      :::: ||| :::: ||| :::: |||
Qy      206 lATHAsnProvalValArgArpNeglyThrAlaThrValAlaGlySerThrValThrs 226
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1305 GGGCCAAAC-----ACAGCCACCACTCTGAGTCCAGTACGACT 1343
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      226 eYAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIlAsnAlaSerAsnAlaGlyT 246
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1344 CCAGTGGG-----GCCAACACAGCCA 1364
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      246 hVAlIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1365 CCAACTGTGATCCAGCACAGTGTCCAGTGGGCCAGCACCTGCCCAACTCTGATCCA 1424
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      264 -----ThrAsnGlyPheValIleSerSerTyrlIleHisIleLeuAspAlaGlnValIleThrs 282
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1425 GCACAACTCTCCAGTGGGGGTGAGCAGACGCCCACTGAGTCCAGCAACTCTCCAGTG 1484
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyAla 302
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1485 GGGCTAGCACAGCCCAACTCTGACTCCAGCACAACTCCAGTGGGCCAGCACAGCCA 1544
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      302 lAlenPheGlnPheArgIleSerArgAsnGlyVal 313
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1545 CCAACTGTGATGTAGCACAGTGTCCAGTGGGATC 1579
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

RESULT 28
US-10-063-502-99
Sequence 99, Application US/10063502
Publication No. US20030023042A1

GENERAL INFORMATION:

APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
CURRENT APPLICATION NUMBER: US/10/063,502
CURRENT FILING DATE: 2002-05-01
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 99
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-063-502-99

Alignment Scores:

Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: 9 Gaps: 12

US-10-018-706-2 (1-322) x US-10-063-502-99 (1-2436)

```

Qy      6 AlaIleAsnSerGlnAsnGlnIleuysProIleAspLeuGlyLeuIlePheGlyValIle 25
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      686 GCCAACAACTCTGATGACGACAGTGTCCAGTGGGCCAGACATGCCCAACTCTGAG 745
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      26 ThrThrCysIleLeuAlaGlyCysAlaSerIleysProThrTyAsnSerThrsSerGlySer 45
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      746 TCTAGCACACTCTCCAGTGGGCC-----ACAGCACCCCAACTCT 787
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      46 GlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThrAsp 65
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

Db      788 GACTTCAGACAACTCTGAGTGGGGGTAGCACAGCCCAACTCTGATCCAGCAACACC 847
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      66 SerGlnGlyValProAsnArgTyrlIn-----ValIleGlnGlyAsp 79
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      848 TCCAGTGGGGCCAG-CACAGCCACCACTCTGAGTCCAGCAAGTGTCCAGTGGGCCAG 906
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      80 ThrValSerIleAlaGlnArgTyrlGlyLeuSerTrpArgGlnIleGlyHis----- 97
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      907 CACTGCCCAACTCTGAGTCCAGCACAACTCCAGTGGGGCCAGCACAGCCCAACTC 966
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      98 -----IleAsnAsnLeu-AsnSerSerTyrlIleTyThrGlyGlnTrpLeuThrle 115
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      967 TGAATCCAGACAGACTCTCCAAATGGGGCTGGGACAGCCCAACTCTGATCCAGCAGAC 1026
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      115 utrpSerGlyAspLeuysValArg-----GlnArgSerIleSerSerGlyValAs 132
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1027 CTCAGTGGGGCCAGCACAGCCCAACTCTGATCCAGCACAGTGTCCAGTGGGGCCAG 1086
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      132 ThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1087 CACTGCCCAACTCTGAGTCCA-----GCCAGCTCTCAATGGGGCCAG 1131
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      150 lngIleProAlaValGlnIleysProThrProProValValValIleIleysProThrP 170
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1132 CACAGCCCAACTCTGAGTCCAGCACAACTCTCCAGTGGGGGTAGCACAGCCCAACTC 1191
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1192 TGACTCCA-----GCACAACTCTCCAGTGGGGCCGACAGCCCAACTCTGATCCAG 1245
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      186 hrgIleAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyA 206
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1246 CACAGTGTCCAGTGGGATGACGACAGTCCCAATTCATGATGCCACACCTCTCC-AGTG 1304
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      206 lATHAsnProvalValArgArpNeglyThrAlaThrValAlaGlySerThrValThrs 226
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1305 GGGCCAAAC-----ACAGCCACCACTCTGAGTCCAGTGGGCCAGT 1343
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      226 eYAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIlAsnAlaSerAsnAlaGlyT 246
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1344 CCAGTGG-----GCCAACACAGCCA 1364
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      246 hVAlIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1365 CCAACTGTGATCCAGCACAGTGTCCAGTGGGCCAGCACCTGCCCAACTCTGATCCA 1424
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      264 -----ThrAsnGlyPheValIleSerSerTyrlIleHisIleLeuAspAlaGlnValIleThrs 282
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1425 GCACAACTCTCCAGTGGGGGTGAGCAGACGCCCACTGAGTCCAGCAACTCTCCAGTG 1484
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyAla 302
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1485 GGGCTAGCACAGCCCAACTCTGACTCCAGCACAACTCCAGTGGGCCAGCACAGCCA 1544
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy      302 lAlenPheGlnPheArgIleSerArgAsnGlyVal 313
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1545 CCAACTGTGATGTAGCACAGTGTCCAGTGGGATC 1579
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

RESULT 29

US-10-173-706-309

Sequence 309, Application US/10173706

Publication No. US20030022293A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.

```

1  APPLICANT:  Wood,William I.
2  APPLICANT:  Zhang,Zemin
3  TITLE OF INVENTION:  SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
4  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
5  FILE REFERENCE:  P343081C7
6  CURRENT APPLICATION NUMBER:  US/10/173,706
7  CURRENT FILING DATE:  2002-06-17
8  Prior Application removed - See File Wrapper or Palm
9  NUMBER OF SEQ ID NOS:  612
10 SEQ ID NO 309
11
12 LENGTH:  2436
13
14 TYPE:  DNA
15
16 ORGANISM:  Homo Sapien
17
18 US-10-173-706-309

```


Db 848 TCACATGGGGCCAG-CACAGCCACCACTGTGATCTCCAGACAGAGTGTCCAGTAGGGCCAG 906
QY 80 ThrValSerIleAlaGlnArgTyrGlyLeuAenTPRArgGluIleGlyHis----- 97
Db 907 CACTGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGACCACTGTGATCTCCAGACCACTC 966
QY 98 -----IleAsnAsnLeu-AsnSerSerTyrThrIleTyrThrGlyIntPLeuThrLe 115
Db 967 TGAGTCCAGAACGACCTCCAAATGGGGCTGGCCAGACCACTGTGATCTCCAGACCACTC 1026
QY 115 uTPSerGlyAspLeuIleValArg-----GluArgSerIleSerSerGlyValAs 132
Db 1027 CTCAGTGGGGCCAGACAGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db 1087 CACTGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCCAG 1131
QY 150 IngIHisProAlaValGlnIleProThrProProValAlaValIleValIleValIleValIle 170
Db 1132 CACAGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCTGGCCAGACCACTC 1191
QY 170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
Db 1192 TGACTCCA-----GCACACCTCCAGTGGGGCCAGACCACTGTGATCTCCAGTGGGGCCAG 1245
QY 186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db 1246 CACAGTGTCCAGTGGGATCCAGACCACTGTGATCTCCAGTGGGGCTGGCCAGACCTCC-AGTG 1304
QY 206 IaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValIle 226
Db 1305 GGGCCCAAC-----ACAGCCACCACTGTGATCTCCAGTGTGATCTCCAGTGTGATCT 1343
QY 226 exAngIlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db 1344 CCAAGTGGG-----GCCAACACAGCCA 1364
QY 246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
Db 1365 CCAACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCCAGACCTCCACCACTGTGATCTCA 1424
QY 264 -----ThrAngIlyPheAlaSerSerTyrIleHisIleIleAspAlaGlnValIle 282
Db 1425 GCACACCTCCAGTGGGGCTGGCCAGACCACTGTGATCTCCAGTGGGGCTGGCCAGT 1484
QY 282 IyAspThrValArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyValAla 302
Db 1485 GGGCTGACACAGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCTGGCCAGT 1544
QY 302 IaLeuPheGluPheAlaGlyIleSerArgAsnGlyVal 313
Db 1545 CCAACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCT 1579

RESULT 31
US-10-175-752-309
Sequence 309, Application US/10175752
Publication No. US2003022295A1
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Matanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C60

US-10-018-706-2 (1-322) x US-10-175-752-309 (1-2436)
CURRENT APPLICATION NUMBER: US/10/175,752
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 309
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-752-309
Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: Gaps: 12
US-10-018-706-2 (1-322) x US-10-175-752-309 (1-2436)
QY 6 AlaIleAsnSerGlnAsnGlnIleProIleValGlyLeuIlePheGlyValIle 25
Db 686 GCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCTGGCCAGACCACTGTGATCT 745
QY 26 ThrThrCylIleLeuAlaGlyCysAlaSerIleProThrTyrAsnSerThrSerGlySer 45
Db 746 TCTAGCACCTCTCCAGTGGGGCC-----AGCACAGCCACCACTCT 787
QY 46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
Db 788 GACTCCAGACCACTCTCCAGTGGGGCTGGCCAGACCACTGTGATCTCCAGACCACTC 847
QY 66 SerGlnIleValProAsnArgTyrGln-----ValGlyGlnIleAsp 79
Db 848 TCCAGTGGGGCCAG-CACAGCCACCACTGTGATCTCCAGTGTGATCTCCAGTGTGATCT 906
QY 80 ThrValSerIleAlaGlnArgTyrGlyLeuAenTPRArgGluIleGlyHis----- 97
Db 907 CACTGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCCAGACCACTGTGATCTC 966
QY 98 -----IleAsnAsnLeu-AsnSerSerTyrThrIleTyrThrGlyIntPLeuThrLe 115
Db 967 TGAGTCCAGAACGACCTCCAAATGGGGCTGGCCAGACCACTGTGATCTCCAGACCACTC 1026
QY 115 uTPSerGlyAspLeuIleValArg-----GluArgSerIleSerSerGlyValAs 132
Db 1027 CTCAGTGGGGCCAGACAGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db 1087 CACTGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCCAG 1131
QY 150 IngIHisProAlaValGlnIleProThrProProValAlaValIleValIleValIleValIle 170
Db 1132 CACAGCCACCACTGTGATCTCCAGACCACTGTGATCTCCAGTGGGGCTGGCCAGACCACTC 1191
QY 170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
Db 1192 TGACTCCA-----GCACACCTCCAGTGGGGCCAGACCACTGTGATCTCCAGTGGGGCCAG 1245
QY 186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db 1246 CACAGTGTCCAGTGGGATCCAGACCACTGTGATCTCCAGTGGGGCTGGCCAGACCTCC-AGTG 1304
QY 206 IaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValIle 226
Db 1305 GGGCCCAAC-----ACAGCCACCACTGTGATCTCCAGTGTGATCTCCAGTGTGATCT 1343
QY 226 exAngIlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db 1344 CCAAGTGGG-----GCCAACACAGCCA 1364
QY 246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263

LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-757-309

Alignment Scores:

	0.00289	Length:	2436
Pred. No.:	122.50	Matches:	78
Score:	39.04%	Conservative:	52
Best Local Similarity:	23.42%	Mismatches:	145
Query Match:	7.34%	Indels:	60
		Gaps:	12

US-10-018-706-2 (1-322) x US-10-176-757-309 (1-2436)

```

QY 6 AAlaIleAsnSerGlnAsnGlnLysProIleYsATGLeuGlyLeuIlePheGlyValIle 25
DB 686 GCACCACTCTGAGTCCAGCAGCACTGTCAGTGGCCAGCAGCCTCCACCACTCTGAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTYrAsnSerThrSerGlySer 45
DB 746 TCTAGCAGCACTCTCCAGTGGGCC-----AGCAGCAGCCACCACTCT 787
QY 46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
DB 788 GACTCCAGCAGCACTCTCCAGTGGGCCCTGACAGCAGCCCACTCTGAGTCCAGCAGCAAC 847
QY 66 SerGlnGlyValProAsnArgTYrGln-----ValIleGlnGlyAsp 79
DB 848 TCCAGTGGGCCAG-CACAGCCACCACTCTGAGTCCAGCAGCAGTGTCCAGTGGGCCAG 906
QY 80 ThrValSerLysIleAlaGlnArgTYrGlyLeuAsnTPRAGluIleGlyHis----- 97
DB 907 CACTGCCACCACTCTGAGTCCAGCAGCACTCTCCAGTGGGCCAGCAGCAGCCCACTCTC 966
QY 98 -----IleAsnAsnLeu-AsnSerSerTYrThrIleTYrThrGlyGlnTPRLeuThrIle 115
DB 967 TGAGTCCAGAGAGCCTCCATGGGGCTGGCAGCAGCAGCACTCTGAGTCCAGCAGCAGC 1026
QY 115 uTPSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
DB 1027 CTCACAGTGGGCCAGCAGCAGCAGCACTCTGAGTCCAGCAGCAGTGTCCAGTGGGCCAG 1066
QY 132 nTPRAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
DB 1087 CACTGCCACCACTCTGAGTCC-----GACAGCCTCCAGTGGGCCAG 1131
QY 150 lngIlnHisProAlaValAlaGlnLysProThrProProValValValLysLysProThrP 170
DB 1132 CACAGCAGCACTCTGAGTCCAGCAGCAGCAGCAGTGTCCAGTGGGCCAGCAGCAGCAGC 1191
QY 170 roTHrProProValValAlaGlnGlnPro--AlaProValAlaProProVal-----T 186
DB 1192 TGAGTCCA-----GCACCACTCTCCAGTGGGCCAGCAGCAGCAGCAGTGTCCAGT 1245
QY 186 hrcIuAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTYrProValGlyA 206
DB 1246 CAGAGTGTCCAGTGGATCCAGCAGCAGTCCCAATTTGATCCAGCAGCAGCCTCC-AGTG 1304
QY 206 lArhAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrS 226
DB 1305 GGGCCCAAC-----ACAGCCCACTCTGAGTCCAGTCCAGTCCAGTCCAGTCCAGT 1343
QY 226 eArgnGlyMetTPRPheserGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
DB 1344 CAGAGTGG-----GCCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1364
QY 246 hrcAlIleGlnAlaAsnHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
DB 1365 CCAAGCTGAGTCCAGCAGTGTCCAGTGGGCCAGCAGTCCAGCAGCAGCAGTGTCCAGTCCA 1424
QY 264 -----ThraGlnGlyPheValSerSerTYrIleHisIleYsAsnAlaGlnValIleYsThng 282

```

```

DB 1425 GCACCACTCTCCAGTGGGCCAGCAGCAGCAGCAGCAGTGTCCAGCAGCAGCAGTGT 1484
QY 282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyValAla 302
DB 1485 GGGCTAGCAGCAGCAGCAGCAGTGTGATCTCCAGCAGCAGCAGTGTCCAGTGGGCCAGCAGCAGCAG 1544
QY 302 lArLeuPheGlnPheArgIleSerArgAsnGlyVal 313
DB 1545 CCAACTGTGAGTCTAGCAGCAGTGTCCAGTGGGATC 1579

```

RESULT 34

US-10-176-913-309
Sequence 309, Application US/10176913
Publication No. US2003002298A1
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Han, James
APPLICANT: Smith, Victoria
APPLICANT: Metanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C66
CURRENT APPLICATION NUMBER: US/10/176,913
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See file wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 309
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-913-309

Alignment Scores:

	0.00289	Length:	2436
Pred. No.:	122.50	Matches:	78
Score:	39.04%	Conservative:	52
Best Local Similarity:	23.42%	Mismatches:	145
Query Match:	7.34%	Indels:	60
		Gaps:	12

US-10-018-706-2 (1-322) x US-10-176-913-309 (1-2436)

```

QY 6 AAlaIleAsnSerGlnAsnGlnLysProIleYsATGLeuGlyLeuIlePheGlyValIle 25
DB 686 GCACCACTCTGAGTCCAGCAGCAGCAGTGTCCAGTGGGCCAGCAGCAGCAGCAGCAGTGTAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTYrAsnSerThrSerGlySer 45
DB 746 TCTAGCAGCACTCTCCAGTGGGCC-----AGCAGCAGCCACCACTCT 787
QY 46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
DB 788 GACTCCAGCAGCACTCTCCAGTGGGCCCTGACAGCAGCAGCAGTGTCCAGTGGGCCAG 847
QY 66 SerGlnGlyValProAsnArgTYrGln-----ValIleGlnGlyAsp 79
DB 848 TCCAGTGGGCCAG-CACAGCCACCACTCTGAGTCCAGCAGCAGCAGTGTCCAGTGGGCCAG 906
QY 80 ThrValSerLysIleAlaGlnArgTYrGlyLeuAsnTPRAGluIleGlyHis----- 97
DB 907 CACTGCCACCACTCTGAGTCCAGCAGCAGCAGCAGTGTCCAGTGGGCCAGCAGCAGCAGCAGC 966
QY 98 -----IleAsnAsnLeu-AsnSerSerTYrThrIleTYrThrGlyGlnTPRLeuThrIle 115
DB 967 TGAGTCCAGAGAGCCTCCATGGGGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1026

```

```

Qy 115 uTrpSerGlyAspLeuysValArg-----GluArgSerIleSerSerGlyValAs 132
Db 1027 CTCAGGTGGGGCGACGACGACCAACTGTGACTCCAGAGTCAGAGTGTCAAGTGGGGCCAG 1086
Qy 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db 1087 CACTGCCACCAACTGTGAGTCA-----GCCAGCACTCCAGTGGGGCCAG 1131
Qy 150 InglnHisProAlaValGlnLysProThrProProValValValValLysLysProThrP 170
Db 1132 CACAGCCACCAACTGTGAGTCCAGACGACACTCTCCAGTGGGGCTTACACAGCCACCACTC 1191
Qy 170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
Db 1192 TGACTCCA-----GCACAACTCCAGTGGGGCCGCGACAGCCAACTCCTGAGTCCAG 1245
Qy 186 hrgGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db 1246 CACAGTGTCCAGTGGGATCCAGACAGTCAACCAATTCTGAGTCCAGACACCTCC-AGTG 1304
Qy 206 laThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrS 226
Db 1305 GGGCCAC-----ACAGCCACCAACTCTGAGTCCAGTCCAGTCCAGTCC 1343
Qy 226 eTAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db 1344 CCAAGTGG-----GCCAACACAGCCA 1364
Qy 246 hTValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
Db 1365 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGACTGCCACCAACTCTGAGTCCA 1424
Qy 264 -----ThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLysThrG 282
Db 1425 GCACAACTTCCAGTGGGGGTTCAGCACAGCCCAACTCTGAGTCCAGCAACCTCCAGTGG 1484
Qy 282 lAspThrValAlaArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaA 302
Db 1485 GGGGTAGACAGACGCCCAACTGTGACTGACGACACCAACTCCAGTGGGCCAGACAGCCA 1544
Qy 302 lAlaLeuPheGluPheArgIleSerArgAsnGlyVal 313
Db 1545 CCAACTGTGAGTCTAGACAGAGTGTCCAGTGGGATC 1579

RESULT 35
US-10-180-552-309
; Sequence 309, Application US/10180552
; Publication No. US20030022300a1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C153
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 309
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-180-552-309

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
6	122.50	2436	78	52	145	60	12
26	39.04%						
Best Local Similarity:	23.42%						
Query Match:	7.34%						

US-10-018-706-2 (1-322) x US-10-180-552-309 (1-2436)

```

Qy 6 AlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeuIlePheGlyValIle 25
Db 686 GCCACCAACTCTGAGTCCAGACAGTGTCCAGTGGGGCCAGCTGCCACCACTCTGAG 745
Qy 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsnSerThrSerGlySer 45
Db 746 TCTGACACACTCTCCAGTGGGGCC-----ACACAGCCACCAACTCT 787
Qy 46 GlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThrAsp 65
Db 788 GACTCCAGACCAACTCTCCAGTGGGGCTTACGACACCACTCTGAGTCCAGACCAACC 847
Qy 66 SerGlnGlyValProAsnArgTyrGln-----ValLysGlnGlyAsp 79
Db 848 TCCAGTGGGGCCAG-CACAGCCACCAACTCTGAGTCCAGACAGTGTCCAGTGGGGCCAG 906
Qy 80 ThrValSerLysIleAlaGlnArgTyrGlyLeuAsnThrPargGluIleGlyHis----- 97
Db 907 CACTGCCACCAACTCTGAGTCCAGACCAACTCTCCAGTGGGGCCAGACAGCCACCACTC 966
Qy 98 -----IleAsnAsnLeu-AsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThrle 115
Db 967 TGACTCCAGAAACACACTCTCAATGGGGCTTACGACAGCCACCACTCTGAGTCCAGACGAC 1026
Qy 115 uTrpSerGlyAspLeuysValArg-----GluArgSerIleSerSerGlyValAs 132
Db 1027 CTCAGTGGGGCCAGACGACCAACTGTGACTCCAGACAGTGTCCAGTGGGGCCAG 1086
Qy 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db 1087 CACTGCCACCAACTGTGAGTCA-----GCCAGCACTCCAGTGGGGCCAG 1131
Qy 150 InglnHisProAlaValGlnLysProThrProProValValValValLysLysProThrP 170
Db 1132 CACAGCCACCAACTGTGAGTCCAGACGACACTCTCCAGTGGGGCTTACACAGCCACCACTC 1191
Qy 170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
Db 1192 TGACTCCA-----GCACAACTCCAGTGGGGCCGCGACAGCCAACTCTGAGTCCAG 1245
Qy 186 hrgGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db 1246 CACAGTGTCCAGTGGGATCCAGACAGTCAACCAATTCTGAGTCCAGACACCTCC-AGTG 1304
Qy 206 laThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrS 226
Db 1305 GGGCCAC-----ACAGCCACCAACTCTGAGTCCAGTCCAGTCCAGTCC 1343
Qy 226 eTAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db 1344 CCAAGTGG-----GCCAACACAGCCA 1364
Qy 246 hTValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
Db 1365 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGACTGCCACCAACTCTGAGTCCA 1424
Qy 264 -----ThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLysThrG 282
Db 1425 GCACAACTTCCAGTGGGGGTTCAGCACAGCCCAACTCTGAGTCCAGCAACCTCCAGTGG 1484
Qy 282 lAspThrValAlaArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaA 302
Db 1485 GGGGTAGACAGACGCCCAACTGTGACTGACGACACCAACTCCAGTGGGCCAGACAGCCA 1544

```

```

QY 302 lAleupheglupheargileserArgasnGlyVal 313
DB 1545 CCAACTCTGAGTCTAGCAGCAGTGTCCAGTGGGATC 1579

RESULT 36
US-10-180-557-309
Sequence 309, Application US/10180557
Publication No. US20030022301A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C147
CURRENT APPLICATION NUMBER: US/10/180,557
CURRENT FILING DATE: 2002-06-25
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 309
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-180-557-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: 9 Gaps: 12

US-10-018-706-2 (1-322) x US-10-180-557-309 (1-2436)
QY 6 AlaIleasnSerGlnasnGlnIysProIleYsArgLeuGlyLeuIleheGlyValIle 25
DB 686 GGCACCACTCTGAGTCTAGCAGCAGTGTCCAGTGGGCGCAGCAGCTGCCCACTCTGAG 745
QY 26 ThrThrcyAlleuAlaGlyCyAlaSerIysProThrIysasnSerThserGlySer 45
DB 746 TCGAGCAGCTCTCCAGTGGGCC-----AGCAGCCCACTCTCT 787
QY 46 GlySerHisArgThserGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
DB 788 GACTCCAGCAGCAACTCTCAGTGGGCTAGCAGCAGCCCACTCTGAGTCCAGCAGCAACC 847
QY 66 SerGlnGlyValProasnArgIysGln-----ValIleGlnIysAsp 79
DB 848 TCCAGTGGGCGCCAG-CACAGCCCACTCTGAGTCCAGCAGCAGTGTCCAGTGGGCGCCAG 906
QY 80 ThrValSerIysIleAlaGlnArgIysGlyLeuAsnTrpArgGlnIleGlyHis----- 97
DB 907 CACTGCCACCACTCTGAGTCCAGCAGCACTCTCAGTGGGCGCAGCAGCAGCCCACTCTC 966
QY 98 -----IleAsnAsnIeu-AsnSerSerIysTrpIleIysTrpGlnIysTrpIle 115
DB 967 TGAGTCCAGAACGACCTCCAGTGGCTGAGCAGCAGCCCACTCTGAGTCCAGCAGCAGC 1026
QY 115 uTpsSerGlyAspLeuIysValArg-----GluArgSerIleSerSerGlyValAs 132
DB 1027 CTCGAGTGGGCGCAGCAGCAGCCCACTCTGAGTCCAGCAGCAGTGTCCAGTGGGCGCCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150

```

```

DB 1087 CACTGCCACCACTCTGAGTCTCA-----GCAGGACCTCCAGTGGGCGCCAG 1131
QY 150 IlnGlnIspProAlaValGlnIysProThrProProValValValIysIysProThrP 170
DB 1132 CACAGCCACCACTCTGAGTCCAGCAGCAGCACTCTCAGTGGGCTAGCAGCAGCCCACTCTC 1191
QY 170 roThrProProValValGlnGlnIspPro-----AlaProValAlaProProVal-----T 186
DB 1192 TGACTCCA-----GCACCACTCTCCAGTGGGCGCCGACAGCCCACTCTGAGTCCAG 1245
QY 186 hrGlnIAspProPheIleThrIysSerSerGlyValIleMetGlnPheArgIysTrpProValGly 206
DB 1246 CACAGTGTCCAGTGGGATCAGCAGCAGTCCCACTCTGAGTCCAGCAGCAGCCCTCC-AGTG 1304
QY 206 lArThrAsnProValValArgAspPheGlyThrAlaThrValAlaGlySerThrValThrS 226
DB 1305 GGGCCAAC-----ACAGCCCACTCTGAGTCCAGTCCAGTCT 1343
QY 226 eArgnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnIleSerAsnIleGlyT 246
DB 1344 CAGTGGG-----GCACACAGCCCA 1364
QY 246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
DB 1365 CCAACTCTGAGTCTAGCAGCAGTGTCCAGTGGGCGCAGCAGCTGCCCACTCTGAGTCCA 1424
QY 264 -----ThrAsnGlyPheValSerSerIysTrpIleHisIleIysAspAlaGlnValIysThrG 282
DB 1425 GCACCAACTCTCAGTGGGCTAGCAGCAGCCCACTCTGAGTCCAGCAGCAACTCTCAGTGTG 1484
QY 282 lYAspThrValArgThrGlnArgIleAlaSerMetIysAsnGlnProSerGlyAla 302
DB 1485 GGGCTAGCAGCAGCCCACTCTGAGTCCAGCAGCACTCTCAGTGGGCGCAGCAGCCCA 1544
QY 302 lAleupheglupheargileserArgasnGlyVal 313
DB 1545 CCAACTCTGAGTCTAGCAGCAGTGTCCAGTGGGATC 1579

RESULT 37
US-10-173-700-309
Sequence 309, Application US/10173700
Publication No. US20030027262A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C14
CURRENT APPLICATION NUMBER: US/10/173,700
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 309
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-173-700-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145

```


Db	1132	CACAGCCCAACCACTGTGAGTCCAGACGACGCTCCAGTGGGGGTAGCACAGCAACCAACTC	1139
Oy	170	rothPrProValValGlnGlnPro--AlaProValAlaProProVal-----T	186
Db	1192	TGACTGCCA-----GCACAACTCCAGTGGGGGGGGCACAGCACCAACTGTGACTCCAG	1245
Oy	186	hrGluAlaProPheAlaThrGlySerSerGlyValaMetGlnPheArgTrpProValGlyA	206
Db	1246	CACAGTGTCCAGTGGGAATCCAGACACCAATTCCTGATGCCAGCAACCTCTCC--AGTG	1304
Oy	206	IaThrAsnProValaValaArgArgPheGlyThrAlaThrValaAlaGlySerThrValThrS	226
Db	1305	GGGGCCAACT-----ACAGCCCAACCACTGTGACTGACGTAGTACACTT	1343
Oy	226	erAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT	246
Db	1344	CCAGTGGG-----GCACACACAGCCA	1364
Oy	246	hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis-----	263
Db	1365	CCAACTGTGAGTCCAGCAACAGTGTCCAGTGGGGCCAGACACTGCCAACCAACTGTGATGCC	1424
Oy	264	-----ThrAsnGlyPheValSerSerTrpIleHisIleIleAspAlaGlnValIlyStrng	282
Db	1425	GCACAACTCCAGTGGGGGTGACGACACGCCAACCAACTGTGAGTCCAGCAACACTCCAGTG	1484
Oy	282	IlyAspThrValArgTrngIyGlnArgIleGlyIleAserMetIlyAsnGlnProSerGlyAla	302
Db	1485	GGGGTACGACAGCAACCAACTGTGACTCCAGACCAACCAACTCCAGTGGGCCACACAGCCA	1544
Oy	302	IaLeuPheGlnPheArgIleSerArgAsnGlyVal	313
Db	1545	CCAACTCTGAGTCTACGACAGTGTCCAGTGGGAATC	1579

```

RESULT 39
US-10-174-579-309
Sequence 309, Application US/10174579
Publication No. US20030027264A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACID
FILE REFERENCE: P3430R1C31
CURRENT APPLICATION NUMBER: US/10/174,579
CURRENT FILING DATE: 2002-06-18
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 309
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-174-579-309

Alignment Scores:
Pred. No.: 0.00289
Score: 122.50
Percent Similarity: 12.50%
Best Local Similarity: 23.04%
Query Match: 7.34%
Gaps: 9
Length: 2436
Matches: 78
Conservative: 52
Mismatch: 145
Indels: 60
Gaps: 12

```

US-10-018-706-2 (1-322) X US-10-174-579-309 (1-2436)

QY	AlaIleasnSerGlnAsnGlnInylProIleLeysArgLeuGlnIleIlePheGlyValIle	25
Db	686 GCCACCACTCTGAGTCACGACCACTGTCACATGAGGCGACACTGCCACCACTCTGAG	745
QY	26 ThrThrCysAlleuAlaGlyCysAlaSerIylProThrIylPheAsnThrSerGlySer	45
Db	746 TCTAGCACTCTCCAGTGGGGCC-----ACGACGCCACCACTCT	787
QY	46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp	65
Db	788 GACTTCAGCAACCTCCAGTGGGGCTAGCAAGCCACCACTCTGATGCCAGCAACC	847
QY	66 SerGlnGlyValProAsnArgIylGln-----ValIylSerGlnIylAsp	79
Db	848 TCCAGTGGGGCCAG--CAGACGCCACCACTCTGAGTCCAGCAAGTGTCCAGTGGGGCCAG	906
QY	80 ThrValSerIylIleAlaGlnArgIylGlyLeuAsnTrpArgGlnIleGlyHis-----	97
Db	907 CACTGCCACCACTCTGAGTCCAGCAACCTCCAGTGGGGCCAGCAAGCCACCACTC	966
QY	98 -----IleAsnAsnLeu--AsnSerSerIylThrIleTylThrGlyGlnTrpPheThrIle	115
Db	967 TGAGTCCCAACAGACTCCATGAGGGGCTGGACAGCACCAACTCTGATGCCAGCGAC	1022
QY	115 uTPSerGlyAspLeuIylValArg-----GlnArgSerIleSerSerGlyValAs	132
Db	1027 CTCACATGGGGGCAGACAGCCACCACTCTGACTCCAGCAAGTGTCCAGTGGGGCCAG	1081
QY	132 nTPAlaIa-HisThrProSerProValAlaValGlnSerSerArgProProVal-----	150
Db	1087 CACTGCCACCACTCTGAGTCCA-----GCACGACCTCCAGTGGGGCCAG	1133
QY	150 InGlnHisProAlaValGlnIylsProThrProProValAlaValValIylValylsProThrP	170
Db	1132 CACAGCCACCACTGAGTCCAGGACGACTCCAGTGGGGCTTAGACAGCCACCACTC	1199
QY	170 rothProProProValValGlnGlnPro---AlaProValAlaProProVal-----	186
Db	1192 TGACTCCA-----GCACACTCCAGTGGGGCGGCGACAGCCACCACTCTGAGTCCAG	1241
QY	186 hncGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgIylProValGlyVal	206
Db	1246 CACAGTGTCCAGTGGGATCAGCAAGTCCCAATCTGAGTCCAGCAAGCACCTTCC--AGTG	1301
QY	206 larThrAspProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrs	226
Db	1305 GGGGCCAAC-----ACAGCCACCACTGAGTCCAGTGGGATC	1341
QY	226 eTPAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT	246
Db	1344 CCAAGTGG-----GCCAACAGCCCA	1361
QY	246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis-----	263
Db	1365 CCAACTGTAGTCCAGCAAGTGTCCAGTGGGGCCAGCACTGCCACCAACTGTGATCCA	1421
QY	264 -----ThrAsnGlyPheValSerSerIylIleHisIleIylAspAlaGlnValIylsThrG	282
Db	1425 GCACCAACTCCAGTGGGTCCAGCACAGCCACCAACTCTGAGTCCAGCACCAACTCCAGG	1481
QY	282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetIylsAsnGlnProSerGlyAla	302
Db	1485 GGGCTTAGCCAGACCCCAACTCTAATCTAGTCCAGCAACCTCCAGTGGAGGCCAGACAGCCA	1544
QY	302 laLeuPheGlnPheArgIleSerArgAsnGlyAla	313
Db	1545 CCAACTGTAGTCTAGCACAGTGTCCAGTGGGATC	1579

```

; Sequence 309, Application US/10174582
; Publication No. US20030027265A1
; GENERAL INFORMATION:

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: July 3, 2003, 18:36:07 ; Search time 1135 Seconds

(without alignments)
4594.667 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670

Sequence: 1 MTVTAINSONOKPIKRLGL.....LFEFRISRNQVYDPLTVLK 322

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Xgapop 10.0, Xgapext 0.5

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh

-Q=/cgn2_1/USFO.spool/US10018706/runat 30062003 091105 23805/app_query.fasta_1.519

-DB=EST -QFMT=fastap -SUFFIX=ret -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45

-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=40 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10018706 @CGN 1 1 156 @runat 30062003 091105 23805 -NCPU=6 -ICPU=3

-NO_MMAR -LARGOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:

1: em_estba:

2: em_esthum:

3: em_estin:

4: em_estmu:

5: em_estov:

6: em_estpl:

7: em_estro:

8: em_hlc:

Result No.	Score	Query Length	DB ID	Description
1	139.5	8.4	784	12 BF866446
2	136	8.1	906	17 CNGS1DTR
3	134	8.0	4812	14 BH770991
4	132	7.9	433	17 BQ791034
5	132	7.9	1096	17 AF094939
6	131	7.8	596	12 BG543924
7	131	7.8	1306	14 BM921488
8	129.5	7.8	398	14 BQ791640
9	127	7.6	890	17 CNGS3DTR
10	126.5	7.6	669	14 BQ704548
11	125.5	7.5	628	10 AV520024
12	125.5	7.5	1279	14 BM911051
13	124.5	7.5	666	12 BG593928
14	124.5	7.5	767	14 BQ505542
15	124.5	7.5	839	12 BF859163
16	124	7.4	285	17 BH705150
17	124	7.4	735	17 AZ187441
18	124	7.4	1252	14 BM913728
19	123.5	7.4	538	10 BE240905
20	123.5	7.4	1002	17 CNGS18ZG
21	123	7.4	689	13 BU220947
22	122.5	7.3	500	9 AU268480
23	122.5	7.3	722	17 AG130154
24	122.5	7.3	1363	17 AG054509
25	122	7.3	450	13 BM109778
26	121.5	7.3	1140	17 AG159588
27	120.5	7.2	508	13 B1174098
28	120.5	7.2	969	17 CNGS3NOR
29	120	7.2	365	9 AL377972
30	120	7.2	511	10 AW586923
31	120	7.2	650	14 BQ147993
32	120	7.2	710	13 B1271940
33	120	7.2	1460	13 BM457264
34	119	7.1	567	12 BG607114
35	119	7.1	909	14 BQ646351
36	118	7.1	443	9 A1484465
37	118	7.1	443	10 BE460175
38	117.5	7.0	615	14 H76194
39	117.5	7.0	754	17 BH244723
40	117	7.0	329	14 U74096
41	117	7.0	586	10 AM033946
42	117	7.0	619	10 AV520316
43	117	7.0	864	17 AZ183812
44	116.5	7.0	717	17 AG132838
45	116.5	7.0	772	13 BM050011

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
BF866446/c
LOCUS
DEFINITION
963069A11.x1 C. reinhardtii CC-1690, Stress condition I, normalized
lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION
BF866446
VERSION
BF866446.1 GI:12256590
KEYWORDS
EST.
SOURCE
Chlamydomonas reinhardtii.
ORGANISM
Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE
1 (bases 1 to 784)
Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.

FEATURES	Source
JOURNAL	Analyses of the Chlamydomonas reinhardtii Genome: A Model, Vascular Plants, project phase 3
COMMENT	Unpublished (2000)
CONTACT	Charles Hauser
DCMB	Box 91000
DURHAM	NC 27708-1000
TEL	919 613 8159
FAX	919 613 8177
EMAIL	chauser@duke.edu
LOCATION/QUALIFIERS	1..784
	/organism="Chlamydomonas reinhardtii"
	/strain="CC-1690 wild type mc+ 21gr"
	/db_xref="taxon:3055"
	/clone_11b="C. reinhardtii CC-1690, stress condition 1, normalized, lambda zap ii"
	/note="vector: plusscript ii SK-, site 1: EcoRI, site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap ii (Stratagene) in the EcoRI (5') and XhoI (3') sites. Plusscript ii SK- plasmids were excised from the lambda zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT	171 a 244 c 196 g 169 t 4 others
ORIGIN	
Alignment Scores:	
Pred. No.:	0.0128
Score:	139.50
Percent Similarity:	40.27%
Best Local Similarity:	26.99%
Query Match:	8.35%
DB:	12
	Gaps: 10
US-10-018-706-2 (1-322) x BF866446 (1-784)	
Qy	62 valilethraspserglnlgyvalprobanarytyrglnvallyglnglnlyaphrval 81
Db	595 ATATATCCCTGAGCGAGCTTGACATCTCTTGACCTGCGCGTGGGAGAAAGATCTATNA 536
Qy	82 Serlysllelalnaglnnarytyrglyleuantrpyagclullleghis----- 97
Db	535 AGTTCCTTACT---CGCTCAATATATGCGGTGGGAGGATTAAGCAAGAAAGCCGGAAG 479
Qy	98 -----IleasnleuanbserSer 104
Db	478 ACCTGCACCTCATCTATGACGAGCTGATATTATGTGTAGACAGAAACAAATATAGAGGGG 419
Qy	105 Tyrthrlerytrthrglyintprleuthleu-----Tyrserglyaspleuhyval 122
Db	418 AACACATGACAAAGAAAGATAGTTACAGCGTCCGTTGCTTACCTCTTGATAGTCNG 359
Qy	123 Arglnharyser-----IleSerSerclValamnthralahisthrProserPro 139
Db	358 CGGAGAGCGGTGGGCGCGCGCGGTTCACCGCGGTACTCGCGGCAATGCGCCAAACACT 299
Qy	140 ValAlaValGlnserSerArgProProValGlnlnlnhIsProAlaValGlnlySProthr 159
Db	298 TCGGTATGAGAGGCGGTGCGCCCGCGT-----CCCCCAAGAGTCGGCGCGCG 248
Qy	160 ProProValValValValylsPProthrProthrProProProValGlnlnlnProAla 179
Db	247 CGCGCGCGT-----CGGCTGCCCGCGCCCGCGT-----CCGAGG 212
Qy	180 Pro-----ValAlaProProValThthGlnAla 188

D _b	211	CCGGCGGCAAGGCCAAGAAAGGCMAAGTAATGCTTGGCCCCCACCAGTTGGGTGC-GGG	153
Oy	189	Pro-----PheAlathrgIysSersegiValmetGlnPheargYyrProvalGly	205
D _b	152	CCGCGCGTGCGCTGCCCGCGTCGGACCTTGGGTGGCATCC-----	111
Oy	206	AlathrasnProvalalValargArgPheglYThrAlthValalagIyseThrValthr	225
D _b	110	-----GGGCGCGCTTGGCCCCCGGCATGTAGCTCTTA	78
Oy	226	SerasnGlyMet-TTrPheSeegiValArgaspGIyaapLeuIIasnnIasernIasi	245
D _b	77	TGTAAACGGCGCTGTCGTCTCAAAAAGGGGGCGCCGGTACNCCAATTCCCTATRGGG	18
Oy	245	yThrValIleGlnala	250
D _b	17	AGTCGTATTCACATCC	2

[illegible]

FEATURES	Location/Qualifiers
source	1. .906
	/organism="Anopheles gambiae"
	/strain="PEST"
	/db_xref="taxon:7165"
	/clone="13P15"
	/clone_1ib="Ncorderame1"
	/note="end : T7"
BASE COUNT	200 a 242 c 248 g 214 t 2 others
ORIGIN	
Alignment Scores:	
Pred. No.:	0.0329 Length: 906
Score:	136.00 Matches: 51
Percent Similarity:	36.74% Conserves: 28
Best Local Similarity:	23.72% Mismatches: 80
Query Match:	8.14% Indels: 57
DB:	17 Gaps: 4
US-10-018-706-2 (1-322) x CNS01JER (1-906)	
QY	105 TTTThrllelYrThrNglyGIntRpleuThrlentRpbSerGlyApLeuLysValArgGlu 122
Db	903 TACACATACGA-GGTGGGACGACACACTGAAGTT---AGCGCGGATGTCGCGCGGCAAA 848

```

Qy 125 ArgSerIleSerSergIyValAsnThrAlaHisThrProSerProValAlaValGlnSer 144
Db 847 AAGTCTCTTGGAGGGGCAAAACCGCCAAAGTACCGCTCTTATCAGTCCCGAATCG 788
Qy 145 SerArgProProValGlnGlnHisProAlaValGlnIlyProThrProProValVal 164
Db 787 TCATGCGCGCGCGTGGACAGCGCTGCTGATTTGGCTCCGACGCGTAAAGTGTCTC--- 731
Qy 165 ValIlyIysProThrProThrProThrProProValValGlnGlnProAlaProValAlaProPro 184
Db 731 ----- 731
Qy 185 ValThrGlnAlaProPheAlaThrGlySerSergIyValMetGlnPheArgTyProVal 204
Db 730 -----GCTCTTACTACTCTCTGAAAGTGC----- 704
Qy 205 GlnAlaThrAsnProValValAlaArgPheGlyThrAlaThrValAlaGlySerThrVal 224
Db 704 ----- 704
Qy 225 ThrSerAnGlyMetTrpPheSergIyArgAspGlyAspLeuIleAsnAlaSerAsnAla 244
Db 703 ---AACAGGATATTGATGATTTCGCTGCGCGGTACCGCGTTATGCTCCGAGCC 647
Qy 245 GlnThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGln 262
Db 646 GGGAAAGGAGTGTTCACGTGCTGACCGAGTGTGCTGCTACGTAACCTGATATGATTAG 587
Qy 263 HisThrAnGlyIlyPheValSerSergIyTyrIleHisIleIlyAspAlaGlnValIlyThrGly 282
Db 586 CATGCGGAAGACTACATCAACGCGCTATGCGCAACAACAGATGCTGTAATTAACGG 527
Qy 283 AspThrValAlaGlnThrGlyGlnArgIleAlaSerMetIlyAsnGlnProSergIyAlaAla 302
Db 526 CAGAACTGCAAGCGCGGAGAAAGATTGCAACATGCGGACGACCGGTACGATTCAGTG 467
Qy 303 LeuPheGlnPheArgIleSerArgAnGlyValTyrValAspPro 317
Db 466 AAGTCGATTTCCAGATCCGCTATAGCGGACGCGCATCGATCCG 422

RESULT 3
BH770991 4812 bp DNA linear GSS 01-MAY-2002
LOCUS LMGtag715 MG1363 Random Sequence Tag Library Lactococcus lactis
DEFINITION subsp. cremoris genomic, DNA sequence.
ACCESSION BH770991
VERSION BH770991.1 GI:20373948
KEYWORDS GSS.
SOURCE Lactococcus lactis subsp. cremoris.
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Lactococcus.
REFERENCE 1 (bases 1 to 4812)
AUTHORS Bojotkin,A., Ehrlich,S.D. and Sorokin,A.
TITLE Studies of genomes of dairy bacteria Lactococcus lactis
JOURNLS Sci. Aliments. (2002) In press
COMMENT Contact: Sorokin A
Geneticque Microbiome
INRA
CRJ INRA, Domaine de Villvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologue in strain H1403 is ycgg (96%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 4784.
location/Qualifiers
1..4812
/organism="Lactococcus lactis subsp. cremoris"
/strain="MG1363"
/db_xref="taxon:1359"
/clone_lib="MG1363 Random Sequence Tag Library"

```

```

/Note="Vector: pSGM02; Site 1: SmaI; Library of
chromosomal fragments of L.lactis strain MG1363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT 1468 a 975 c 800 g 1569 t
ORIGIN
Alignment Scores:
Pred. No.: 0.551 Length: 4812
Score: 134.00 Matches: 74
Percent Similarity: 37.37% Conservative: 37
Best Local Similarity: 24.92% Mismatches: 97
Query Match: 8.02% Indels: 89
DB: 17 Gaps: 13
US-10-018-706-2 (1-322) x BH770991 (1-4812)
Qy 26 ThrThrCysIleLeuAlaGlyCysAlaSerIlyProThrTyrAsnSerThr----- 42
Db 102 ACTGCTGCTTGGACGGGAGATATGCTACCGACCAAGTATGCGCTTCACTGAAATACGAT 161
Qy 43 -----SergIySergIlySerHisArgThrSergIy 52
Db 162 ATTATTTCTCAATATTAATTGACTGCTTTGACGGAGCTTCTTCACTGAAATACGAT 221
Qy 53 SerGlyGlyLeuAlaIleGlySergIlyValIleThrAspSergIlyValProAsn--- 71
Db 222 TCTGTGTGC-----TGACACACCAATTTACGATATTAATTTGGAACCAATACG 272
Qy 72 -----ArgTyrGlnValIlyGlnGlyAspThrValSerIlyIleAlaGlnArgTyr 88
Db 273 AGTTCAACTACTATATACCGTCAATCTGTGATCTCTTTGGGAAATCTCAAAAGATAT 332
Qy 89 GlnLeuAsnTrpArgGlnIleGlyHisIleAsnAsnLeuAsnSerTyrThrIleTyr 108
Db 333 GGAATTATGTGCTGCTCAAAATTCAAAGTGCAGATTAATTTAAAGTACC---ATTATCTAC 389
Qy 109 ThrGlyGlnTrpLeuThrLeuTrpSergIyAspLeuIlyValArgGluArgSerIleSer 128
Db 390 ATTGTCGAAAACTGTGACTGACAGCTTCA-----GCTTCTTCTCAAAATTC 437
Qy 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSergArgProPro 148
Db 438 GGTGTTTCAAACAATTC-----GCACACACTACCTCCAAACC 473
Qy 149 ValGlnGlnHisProAlaValGlnIlyProThrProProValValIlyValIlyProPro 168
Db 474 ACTTCTGTGACACCTGCT-----AAACCAACTTCACAAACACTGTTAAGAGGAAATCC 527
Qy 169 ThrProThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAla 188
Db 528 GGAGATACCTTGGGCGGCTATGCTGTAATTAATAACTAGTA----- 569
Qy 189 ProPheAlaThrGlySerSergIyValMetGlnPheArgTyrProValGly-AlaThrAs 208
Db 570 ---GTGCTCATGAAAGTGCAGATCATCTAAGTTACAGATCACTTATATTTGCTCAAAA 626
Qy 208 nProValValArgArgPheGlyThrAla-----ThrValAlaGlySerThrVal 224
Db 627 TCTTATTTTTCAAATCTGCTGCTCTCAAAATCTTCGACAGGTTCAAGCTTCACATGTC 686
Qy 224 lThrSerAnGlyMetTrpPheSergIyArgAspGlyAspLeuIleAsnAlaSerAsnAl 244
Db 687 TACCAATAC-----TCAAACTC 704
Qy 244 aglyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHisThr 264
Db 705 G-----AC 707
Qy 264 rAnGlyIlyPheValSerSergIyTyrIleHisIleIlyAspAlaGlnValIlyThrGlyAspTh 284
Db 708 TTTCTTCACTCAAAATGCTCAATTCAT-----AAGTCTGTTAAAGAGATAC 755
Qy 284 rVal-----ArgThrGlyGlnArgIleAlaSerMetIlyS 295

```

Db 756 TCCTGTGGGACCTTTCGCAAAATCTGGCAGCCCAATTCCTTCATCAAG 804
RESULT 4
LOCUS BQ791034
DEFINITION BQ791034 433 bp mRNA linear EST 30-JUL-2002
E3832 Chinese cabbage etiolated seedling library Brassica rapa
subsp. pekinensis cDNA clone E3832, mRNA sequence.
ACCESSION BQ791034
VERSION BQ791034.1 GI:22005996
KEYWORDS EST
SOURCE Brassica rapa subsp. pekinensis.
ORGANISM Brassica rapa subsp. pekinensis
Bakaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 433)
AUTHORS Kyu,S.H., Yang,K.A., Lee,S.Y., Kim,H.-I., Cho,M.J. and Lim,C.O.
TITLE Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
(2002)
JOURNAL Unpublished (2002)
COMMENT Contact: Lim, C.O.
Plant Molecular Biology & Biotechnology Research Centre
Gyeongsang National University
#900 Gazwa-dong, Jinju 660-701, Korea
Tel: 82 55 751 6255
Fax: 82 55 759 9363
Email: collim@nongae.gsnu.ac.kr
Seq primer: T7.
Location/Qualifiers
source 1..433
/organism="Brassica rapa subsp. pekinensis"
/cultivar="Jangwon"
/db_xref="taxon:51351"
/clone_1b="Chinese cabbage etiolated seedling library"
/issue_type="Etiolated seedling"
/lab_host="Xt-1 Blue"
/note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"
BASE COUNT 125 a 209 c 50 g 49 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0263 Length: 433
Score: 132.00 Matches: 35
Percent Similarity: 41.86% Conservative: 1
Best Local Similarity: 40.70% Mismatches: 38
Query Match: 7.90% Indels: 12
Gaps: 2
US-10-018-706-2 (1-322) x BQ791034 (1-433)
Qy 128 SerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgPro 147
Db 87 AGTAGTGGCCACCGCCACACCAACCAACCCACCTGTGTAAAGCCACCAACACCA 146
Qy 148 ProValGlnGlnHisProAlaValGlnLysProThrProProValAlaValAlaLysLys 167
Db 147 CCAACACCACTCCACAGTCGTAAAGCCACCAACACAGCCCACTGTGTAAAGCCA 206
Qy 168 ProThrProThrProProValAlaGlnGlnProAlaProValAlaProProValThr--- 186
Db 207 CCAACACCAACCAACCGGTGTAAAGCCACCAACCAACCCCTCTGTGTAAACCT 266
Qy 187 ---GluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGly 205
Db 267 CCAACACCAACCGGTGTAAAG---CCACCGACA 296
Qy 206 AlaThrAsnProValAla 211
Db 297 CCAACTCTCTCTGTGTAA 314
RESULT 5

AF094939/c
LOCUS AF094939 1096 bp DNA linear GSS 29-AUG-2000
DEFINITION AF094939 Salmonella typhimurium LT2, Lambda DASH II Salmonella
typhimurium genomic clone 226-T3, DNA sequence.
ACCESSION AF094939
VERSION AF094939.1 GI:4322781
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella
REFERENCE 1 (bases 1 to 1096)
AUTHORS Mong,R.M.Y. and McClelland,M.
TITLE End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,
LI-Cor
JOURNAL Unpublished (1999)
COMMENT Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@itsci.sdu.edu
Class: shotgun.
Location/Qualifiers
source 1..1096
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/clone_1b="226-T3"
/clone_1b="Salmonella typhimurium LT2, Lambda DASH II"
/note="Vector: Lambda DASH II; sequenced using LI-Cor
sequencer"
BASE COUNT 239 a 323 c 272 g 262 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0998 Length: 1096
Score: 132.00 Matches: 85
Percent Similarity: 34.06% Conservative: 40
Best Local Similarity: 23.16% Mismatches: 134
Query Match: 7.90% Indels: 108
Gaps: 14
US-10-018-706-2 (1-322) x AF094939 (1-1096)
Qy 42 ThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerG 61
Db 1059 ACCGAGGCGCCAGCAACCAATCGATCGAGCCGCGAGAGATGAAGTGAATTCCTCAAGA 1000
Qy 61 nValIleThrAspSerGln-----GlyValProAsnArgTyrGlnValLysG 77
Db 999 TGAAGCTGACGATTAACCAAGCGAGCGAGTCGGCGC--CATGAATACGTCGTCCAC 943
Qy 77 nGlyAAPTThrValSerLysIleAlaGlnArgTyrGlyLeuAsnThrPArgGlnIleGlyH 97
Db 942 AGCGATAGCATGTAAAGCAGATTCGTAGATCAGTACGCAATCGATAGCAGATATTAGCG 883
Qy 97 s-----IleAsnAsnLeuAsnSerSerTyrThrIleTyrThcI 110
Db 882 ACTTGGCGCTTCGATTAAGAGACGTGCGCAATGTGAATTTGGCCCAACGCTT----- 831
Qy 110 yGlnTPrPLeuThrLeuTPrSerGlyAspLeuLysValArgLysGserIleSerSerG 130
Db 830 -TCTGTACACATGACCGCTGATGAGCATTTAAAGCGGTGTACATGGAGAAAGTCCCGCG 772
Qy 130 yValAsnThrAlaHisThrProSer-ProVal-----AlaValGlnSerSerA 146
Db 771 TGAACGCGTACCTAGATCGACATGCGCAACGGTTTAAATGACAGTGAATGACACA 712
Qy 146 rGProPro-----ValGlnGlnHisProAlaV 155
Db 711 GGGGAGACTGGGTCAACAGTCTGTAAGAGTACGATAGGAGGTAGCTTGTGCGCAGCGC 652
Qy 155 aGlnLysProThrProProValAlaValLys-----LysProThrPro----- 170

```

Db      651 GAAAGAGCCCGCTTTAAACGACGGAATATCAGCGAGTGAATAAAGCCATGACAGTGCA 592
      170 -----
Db      591 GATGATTTTGGCAAGCTGAAAAAGGCGAGTAATTTTCGGTTCTGATGTCGGCGAGAT 532
      171 -----
Db      531 GCTGATGCAAGGCGTGAACAGAGTCACTGTTGGGCGTGGGAGATGCGTTCCGATGTGA 472
      176 TnglnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySer-Ser 195
      471 AGATTACTACCCGCTTCGCGCGGTGACGTAATTCATGACCGTATGAGGTTGCTCT 412
      196 GlnValMetGlnPhe-ArgTyrProValGly-----
      411 GCGCAAGAGCTTTTACGCTTCGCGACCGCTAAACAGTTCGCGATCTCTCCAACTGCA 352
      206 -----AlaThrAsnProValAlaArgArgPheGlyThr-----Al 217
      351 TCCGCGTGTCTGAACCGGTTACCGGACGCGTTCGCGCGATGCGCTGACTTGC 292
      217 aThrValAlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAs 237
      291 GATGCCGCGAGGTACGCGCGTGTCTGCGTGGG-----GATGGCGA 250
      237 PheAlaLeuAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAl 257
      249 GGTGCGTGTCTGAACCGTACCGCGCGCTGCC-----GGTTA 214
      257 aSerIleValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleValAspAl 277
      213 CTACATTCGCGATTTGTATGACGACCACTACACACAGCTTACATGCACTTCGCTAAGCT 154
      277 aGlnValIleThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsn-- 296
      153 GCTGCTGAACCGGCGCAAAAGTGAACGTGCGATGTATGCGCTTCTGTGTAACAC 94
      297 ----GlnProSerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrVa 315
      93 CGGCGCTTCACAGCGCGCATCTG--CAATTATGAGTATGATCAACGCAAGCCGCT 37
      315 lAspProLeuThr 319
      36 TAACCTCTGACA 24

RESULT 6
LOCUS      BG543924      596 bp      mRNA      linear      EST 01-MAY-2002
DEFINITION El673 Chinese cabbage etiolated seedling library Brassica rapa
ACCESSION  BG543924
VERSION     BG543924.1  GI:20374904
KEYWORDS
SOURCE
ORGANISM   Brassica rapa subsp. pekinensis.
            Bukaraota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
            1 (bases 1 to 596)
            Ryu,S.H., Kang,J.S., Kang,C.-h., Kim,C.Y., Choi,Y.J., Lee,S.-H.,
            Bahk,J.D., Lee,S.Y., Cho,M.J. and Lim,C.O.
            Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
            Unpublished (2001)
            Contact: Lim, C.O.
            Plant Molecular Biology & Biotechnology Research Centre
            Gyeongsang National University
            #900 Gajwa-dong, Jinju 660-701, Korea
            Tel: 82 55 751 6255
            Fax: 82 55 759 9363
            Email: colim@nongae.gsnu.ac.kr
            Seq primer: T7.
  
```

```

FEATURES
  source
    location/Qualifiers
      1..596
      /organism="Brassica rapa subsp. pekinensis"
      /cultivar="Jangwon"
      /db_xref="taxon:51351"
      /clone="E1673"
      /clone_1db="Chinese cabbage etiolated seedling library"
      /issue_type="Etiolated seedling"
      /lab_host="XL-1 Blue"
      /note="vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"
  BASE COUNT      154 a      229 c      100 g      113 t
  ORIGIN
  Alignment Scores:
    Pred. No.:      0.0512      Length:      596
    Score:          131.00      Matches:      33
    Percent Similarity: 43.59%      Conservative: 1
    Best Local Similarity: 42.31%      Mismatches: 32
    Query Match:      7.84%      Indels:      12
    DB:               12      Gaps:      2
  US-10-018-706-2 (1-322) x BG543924 (1-596)
  QY      136 ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal 155
      32 ACACCAACCCCACTGTTGTATAGCGCACCAACACCAACCAACCACTCCACGCTGCTT 91
  Db
  QY      156 GlnLysProThrProProValAlaValAlaValLysProThrProThrProProValAl 175
      92 ACGCCACCAACACACAGCCCACTGTTGTATACGCCACCAACACCAACACCAACCGGTGTA 151
  Db
  QY      176 GlnGlnProAlaProValAlaProProValThr-----GlnAlaProPheAlaThrGly 193
      152 ACGCCACGACACCAACCCCTCCTGCTGTAACCTCAACACCAACCGGTGTAAGC--- 208
  Db
  QY      194 SerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValAl 211
      209 -----CAACGACACCAACTCCTCCTGTCGTA 235
  Db

RESULT 7
LOCUS      BM921488      1306 bp      mRNA      linear      EST 12-MAR-2002
DEFINITION AGENCOURT 6626294 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5752955
ACCESSION  BM921488
VERSION     BM921488.1  GI:19371867
KEYWORDS
SOURCE
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
            1 (bases 1 to 1306)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: egadps-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
            DNA Sequencing by: Agencourt Bioscience Corporation
            clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: LRAM2787 row: p column: 12
            High quality sequence start: 90
            High quality sequence stop: 471.
            Location/Qualifiers
              1..1306
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:5752955"
              /clone_1db="NIH_MGC_115"
  
```


REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
1 (bases 1 to 890)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brotier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 890)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billaud,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Unpublished
3 (bases 1 to 890)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.
Location/Qualifiers
1..890
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="019021"
/clone_id="G"
/note="Genoscope sequence ID : COBG019AH11P1-end : T7"
BASE COUNT 172 a 243 c 338 g 128 t 9 others
ORIGIN

Alignment Scores:
Pred. No.: 0.21 Length: 890
Score: 127.00 Matches: 60
Percent Similarity: 36.73% Conservative: 23
Best Local Similarity: 26.55% Mismatches: 83
Query Match: 7.60% Indels: 60
DB: 17 Gaps: 9

US-10-018-706-2 (1-322) x CNS03DYV (1-890)

25 ILeThrThrCysIleLeuAlaGlyCysAlaSerIysProThrThyRanserThrSergly 44
Db 765 TTGACCTCTGCGACGCTCTCGGCTGAGCGGAGAGCTGGCAGCATCTGAGCGACGCTCC 706
Qy 45 Ser-----GlySerH1sArgThrSerGlySerglyLeuAlaIleGlySerglnVal 62
Db 705 GAGTCTTTTGAGMAAAGCCACCGTGGAAAC-----ATG 670
Qy 63 ILeThrAspSerGlnGlyValProAsnArgTyrlinVallySerglnGlyAspThr----- 80
Db 669 TACCGCAGCGCGCGGTACAGTGTGTTTGAATTAACCCCTGAGTACACCCCGTTT 610
Qy 81 -----ValSerIysIleAlaGlnArgTyrlGlyLeuAsnTrpArgGlnIleGlyHisIle 98
Db 609 CACCTTCGAGAGAAAACGAGAGAGTGGGTTCAAC----- 571
Qy 99 AsnAsnLeuAsnSerIysTyrlIleTyrlThrlGlyGlnTrpLeuThrLeuTrpSergly 118
Db 570 -----AGCTGACCTGTTTGAACAAGCAACCCCTAACATGCTGG----- 532
Qy 119 AspleuLysValArgGluArgSerIleSerSerglyValAsnThrAlaHisThrProSer 138
Db 531 -----TTTCAGAGGTACCCCTCCCTCCCTTGGGGGCTCT 499
Qy 139 ProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal----- 155
Db 498 CCGTGGGGCTGTCAGTCCCGACAGTCCCGCACTCTCCCGCACTCCCGCACTCTCCCACT 439
Qy 156 -----GlnLysProThrProProValAlaValAlaValLys 167

Db 438 CCCCCCAGCCCCCAGTCTCCCGACGTCGCCCGCCAGTCCCGCCAGTCTCC 379
Qy 168 ProThr-----ProThrProProValAlaGlnGlnProAlaPro-ValAlaProProValTh 186
Db 378 CCAATCCCCCCCCCTCCCTCC-----CAGTCCCCAGTCTCTCCCGCCAGTCTCT 325
Qy 186 rGluAlaProPheAlaThrGlySerSerglyValMetGlnPheArgTyrlProValGlyAl 206
Db 324 CCCCAGTCCCCCAGTCCCGCCAGTCTCT-----CCAGTCCCCCA 286
Qy 206 arhAsnProValValArgArgPheGlyThrAla-ThrValAlaGlySerThrValThrs 226
Db 285 GTCTCTCCCGACGTCGCCCGACGTCCTCAACACAGTGAAGCGGAGGCTTAACCCGCCA 226
Qy 226 eArgnGlyMetTrp 230
Db 225 GCACTGGTCTGTGG 212

RESULT 10
B0704548
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

B0704548 669 bp mRNA linear EST 16-JUL-2002
Bn01_03006_A
Bn01_AAFc_ECORC transgenic Brassica napus overexpressing BNCBP17, co
nstitutively frost_tolerant Brassica napus cDNA clone Bn01_03006,
mRNA sequence.
B0704548
B0704548.1 GI:21843967
EST.
rape.
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 669)
Singh,J., Allard,G., Tinker,N., Robert,L., Lacroix,C., De Moors,A.,
Chagnon,J., Farah,S., Couroux,P. and Hattori,J.
Expressed Sequence Tags from constitutively frost tolerant
transgenic Brassica napus overexpressing BNCBP17
Unpublished (2002)
Contact: Singh,J.A.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-Food Canada
KW Neaby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A
0C6, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhja@em.agr.ca.
Location/Qualifiers
1..669
/organism="Brassica napus"
/cultivar="Westar"
/db_xref="taxon:3708"
/clone="Bn01_03006"
/clone_id="Bn01_03006"
/tissue_type="fourth leaf"
/dev_stage="3 weeks seedling grown at room temperature"
/note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
Site 2: XhoI; Germinated in soil flats and seedlings grown
for 3 weeks in a Conviron E-15 cabinet set at 20°C/16 hr
light (250 Em-2sec-1) and 16°C/8 hr dark. Fourth leaves
collected at 9 am and immediately frozen."
BASE COUNT 167 a 203 c 124 g 168 t 7 others
ORIGIN

Alignment Scores:
Pred. No.: 0.155 Length: 669
Score: 126.50 Matches: 32
Percent Similarity: 48.57% Conservative: 2
Best Local Similarity: 45.71% Mismatches: 25
Query Match: 7.57% Indels: 11
DB: 14 Gaps: 3

US-10-018-706-2 (1-322) x BQ704548 (1-669)

QY 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
 DB 35 GTGTAACCCGACCAACCAACCCCTCTGTCGTAAACCAACCAACCAACCGATCGTA 94
 QY 151 GlnHisProAlaValGlnIlyProThrProProValAla-----Val 164
 DB 95 AGCCGACCGACA-----CCACCCCTCTGTCGTAAACCAACCAACCAACCGATC 145
 QY 165 ValIlyLysProThrProThrProProValAlaGlnGlnProAlaPro-----ValAla 182
 DB 146 ATACCCGACCGACCAACCAACCCCTCTGTCGTAAACCAACCAACCGATCGTAACG 205
 QY 183 ProProValThrGlnAlaProPheAlaThr 192
 DB 206 CCACGACACCAACCCCTCTGTCGTAAACCA 235

RESULT 11
 AVS20024/c 628 bp mRNA linear EST 06-SEP-2000
 LOCUS AVS20024 Arabidopsis thaliana aboveground organs two to six-week
 DEFINITION AVS20024 Arabidopsis thaliana cDNA AP204g09F 3', mRNA sequence.
 ACCESSION AVS20024
 VERSION AVS20024.1 GI:8679551
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
 1 (bases 1 to 628)
 REFERENCE Asamizu E, Nakamura Y, Sato S, and Tabata S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7, 175-180 (2000)
 JOURNAL 20363093
 MEDLINE
 COMMENT Contact: Erika Asamizu
 The First Laboratory For Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 232-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1..628
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="AP204g09F"
 /clone_1lb="Arabidopsis thaliana aboveground organs two to
 six-week old"
 /tissue_type="aboveground organs"
 /dev_stage="two to six-week old"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 157 a 105 c 197 g 169 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.174 Length: 628
 Score: 125.50 Matches: 31
 Percent Similarity: 46.25% Conservative: 6
 Best Local Similarity: 38.75% Mismatches: 32
 Query Match: 7.51% Indels: 11
 DB: 10 Gaps: 3

US-10-018-706-2 (1-322) x AVS20024 (1-628)

QY 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
 DB 609 GTATATATCTCCACCAACCAACCAACCCGTCGTGTA-----ACACCGCCCAACGCGCA 559

QY 151 GlnHisProAlaValGlnIlyProThrProProValAlaValIlyLysProThrPro 170
 DB 558 ACTCCACCCGTCATATTCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 499
 QY 171 ThrProProValAlaGlnGlnProAlaProProValAlaProProValThrGlnAlaProPhe 190
 DB 498 ACACCAACCCGTCGTGACCTCCACCAACCAACCAACCAACCAACCAACCAACCAACCA 442
 QY 191 AlaThrGlySerSerGlyValMetGlnPheArgTyrProValAlaGlyAlaThrAspProVal 210
 DB 441 CCAATCTCCCTCCTACG-----CCTATTCAGACACTTGCCTCAATT 403

RESULT 12
 BM911051 1279 bp mRNA linear EST 12-MAR-2002
 LOCUS BM911051
 DEFINITION AGENCOURT 6613368 NIH_MGC_98 Homo sapiens cDNA IMAGE:5455034
 5', mRNA sequence.
 ACCESSION BM911051
 VERSION BM911051.1 GI:19361430
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
 1 (bases 1 to 1279)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: LHCN951 row: c column: 03
 High quality sequence start: 6
 High quality sequence stop: 259.
 Location/Qualifiers

FEATURES
 source
 1..1279
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5455034"
 /clone_1lb="NIH_MGC_98"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

BASE COUNT 332 a 531 c 231 g 184 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.485 Length: 1279
 Score: 125.50 Matches: 51
 Percent Similarity: 36.83% Conservative: 22
 Best Local Similarity: 27.13% Mismatches: 66
 Query Match: 7.51% Indels: 50
 DB: 14 Gaps: 7

US-10-018-706-2 (1-322) x BM911051 (1-1279)

QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyValLeuAlaIleGlySer 60
 DB 100 TCTCTCGGCTGCTCAGGTCCCAACCGCTCCG-GGCACGAGAGGT---GCTTCCGCGCGG 155

Qy 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnVal-----Lys 76
 Db 156 AATGTGACTGAGTCTGGAGAGCCGAGTGGTGAAGGATGACCAAGGGGGGGGCGC 215
 Qy 77 GlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTyrArgGlnIleGly 96
 Db 216 CTTGAAATATGCGCGCTTGTCTTCGAGCGGGGAGACACCGGGGGGGAACAGGGA 275
 Qy 97 HisIle-----AsnAsnLeuAsnSer 103
 Db 276 GGAATCTCCGAAACAAATGCTCGCGCTGCCATCCGGGTACGAATTCATGGAACCGA 335
 Qy 104 SerTyrThrIleTyrGlnGlnTyrLeuThrIleTyrSerGlyAspLeuValArg 123
 Db 336 ACCCAGCACCATTACAGACAAATCTTAACGTTA----- 371
 Qy 124 GluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGln 143
 Db 372 -----CCCAACACCCACACACCCCGCAGCAGCACCACCCCGCCCACTACCA 422
 Qy 144 SerSerArgProProValGlnGlnHisProAlaValGlnLysProThrProProValVal 163
 Db 423 TACCCCAACCTCCCTC---CACCACTACTCTCATCACCCTCCCACTC----- 470
 Qy 164 ValValLysLysProThrProThrProProValValGlnGlnProAlaProValAlaPro 183
 Db 471 -----AAACCCAGCGCCCGCCACACACCAAAAGAGCGCGACCGCAACCGAACCC 521
 Qy 184 ProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrPro 203
 Db 522 GCCGTAAACGACGAGCT-----CCT 542
 Qy 204 ValGlyAlaThrAsnProValVal 211
 Db 543 ACGGCCAGCATTAACCGATCATC 566

RESULT 13
 BG593928 666 bp mRNA linear EST 12-APR-2001
 LOCUS EST12957606 csts Solanum tuberosum cDNA clone csts516 5' sequence,
 DEFINITION mRNA sequence.
 ACCESSION BG593928
 VERSION BG593928.1 GI:13612068
 KEYWORDS EST.
 SOURCE Solanum tuberosum
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 666)
 van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,
 Bougri,O., Buell,C.R., Romning,C., Tanksey,S. and Baker,B.
 Generations of ESTs from sprouting potato eyes
 Unpublished (2000)
 TITLE CONTACT: Cathy Romning
 JOURNAL The Institute for Genomic Research
 COMMENT For clone info: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cdna@resgen.com
 Seq primer: M13F-R.
 FEATURES
 SOURCE Location/Qualifiers
 1..666
 /organism="Solanum tuberosum"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="csts516"
 /clone_lib="csts"
 /label="csts"
 /label_type="sprouting eyes from tubers"
 /dev_stage="12-14 weeks post harvest"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
 taken from tubers. The tubers were incubated at 26C in the
 dark for 2-3 weeks prior to sprouting. The eyes were

BASE COUNT 139 a 212 c 134 g 181 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.234 Length: 666
 Score: 124.50 Matches: 49
 Percent Similarity: 35.96% Conservative: 15
 Best local Similarity: 27.53% Mismatches: 65
 Query Match: 7.46% Indels: 50
 DB: 12 Gaps: 6
 US-10-018-706-2 (1-322) x BG593928 (1-666)

Qy 115 LeuThrSerGlyAspLeuLysValArgGluArgSerIleSerSerGlyValAsn----- 132
 Db 80 CTTGGATCGGGTTTCTGTCATT-----ATCGTGGTGGTGAAGGGGTCAA 127
 Qy 133 -----ThrAlaHisThrProSerPro 139
 Db 128 GCTCTGCAAGTCAACCCACCACTAGTCCCGCAGCACCACCACTACCTACCTGCACT 187
 Qy 140 Val-----AlaValGlnSerSerArgProProValGlnGlnHisProAla 154
 Db 188 ACCGGTTCTCCACCTCTCGAGTAACTTCACTCTCATCTGTATCTCTCCACCGGCA 247
 Qy 155 ValGlnLysProThrProProValValValValLysLysProThrProThrProProVal 174
 Db 248 ACTTCAACCACTAGTACTCTTCCACCAAGTGGCTCCACCACTAGTGGCTTCCACCA 307
 Qy 175 ValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySer 194
 Db 308 GCTACTCTTCCACAGAGAGTGTCTCTTCACTGTGATCTCAAC-TCACCGCGAGC 366
 Qy 195 SerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValAlaArgArgPhe 214
 Db 366 ----- 366
 Qy 215 GlyThrAlaThrValAlaGlySerThrValThrSerAsnLysMetTyrPheSerGlyArg 234
 Db 367 GGCACCTGACCGGTCGCAACTCCTGCTTCACTCC---TGCTCTCTCTCGACTAC 423
 Qy 235 AspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaPheIleAsnMet 254
 Db 424 AAAGTAGCTACATCTCCGACCTTCCCGGAGTGTGATGCTTCCAGACCAACC 483
 Qy 255 AspGlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSerTyrIle 272
 Db 484 TCTGGGTGCTCCGTC-----ACCGAGTGTATTATCTCTCTCATC 525

RESULT 14
 BQ505542 767 bp mRNA linear EST 22-JUN-2002
 LOCUS EST612957 Generation of a set of potato cDNA clones for microarray
 DEFINITION analyses mixed potato tissues Solanum tuberosum cDNA clone STM6E76
 5' end, mRNA sequence.
 ACCESSION BQ505542
 VERSION BQ505542.2 GI:21921472
 KEYWORDS EST.
 SOURCE Solanum tuberosum
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 767)
 Buell,C.R., Hart,A., Baker,B., Tanksey,S., Fry,W., Smart,C.,
 Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
 Karamycheva,S.A.
 Generation of a set of potato cDNA clones for microarray analyses
 Unpublished (2002)
 TITLE CONTACT: Cathy Romning
 JOURNAL The Institute for Genomic Research
 COMMENT On Jun 10, 2002 this sequence version replaced gi:21364411.
 Other ESTs: EST612958

Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potatob@tigr.org
This clone is available through the Research Genetics, contact the
Research Genetics for further information 1-800-711-6195 or
cdna@resgen.com
Seq primer: 13

FEATURES
source location/Qualifiers
1..767
/organism="Solanum tuberosum"
/cultivar="Kennebec or Bintje"
/db_xref="taxon:4113"
/clone_lib="STMBER76"
/clone_1ib="Generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
tubers, or roots."

BASE COUNT 164 a 225 c 152 g 226 t
ORIGIN

Alignment Scores:

Pred. No.: 0.287 Length: 767
Score: 124.50 Matches: 49
Percent Similarity: 35.96% Conservative: 15
Best Local Similarity: 27.53% Mismatches: 65
Query Match: 7.46% Indels: 50
Gaps: 6

US-10-018-706-2 (1-322) x BQ505542 (1-767)

```

QY 115 leuTserGlyAspLeuYsValArgGluArgSerLysSerSerGlyValAsn----- 132
DB 80 CTTTGATCGGGGTTTCTTCGCACT-----ATGTCGCTGGGTAGGGGGTCA 127
QY 133 -----ThAlAHisthProSerPro 139
DB 128 GCTCTGCGCACTGACCACTAGTCCGCGACCACTCACTACCTCCCTGACCT 187
QY 140 Val-----AlaValGlnSerSerArgProProValGlnGlnHisProAla 154
DB 188 ACCGTTCTCCACTCTCTGCAAGTCACTCCCTCCATCTGTCTCTCCACCGGCA 247
QY 155 ValGlnLysProThrProProValValValValLysProThrProThrProProVal 174
DB 248 ACTTACACCACTCTCTCTCCACAGAGAGCGGCTCCACAGTGTCTCCACACCA 307
QY 175 ValGlnGlnProAlaProValAlaProProValThGlnAlaProPhelAlaThGlySer 194
DB 308 GCTACCTCTCCACAGAGAGTGTCTCTCCACCTGTGAGTACCTCACCTCCACCGGCA 366
QY 195 SerGlyValMetGlnHeaArgTyProValGlnAlaThAsnProValAlaArgArgPhe 214
DB 366 ----- 366
QY 215 GlyThrAlaThrValAlaGlySerThrValThSerAsnGlyMetTyrPheSerGlyArg 234
DB 367 GGGACCTGGACCGGTGCAACTGCTGCTTCACTGCTCC--TGCCTTACTCTCCACTAC 423
QY 235 AspGlyAspLeuIleAsnAlaSerAsnAlaGlyThValIleGlnAlaAspHisAsnMet 254
DB 424 AAGAGTAGCTACATCTCTGGCGGCTTGGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAG 483
QY 255 AspGlyAlaSerIleValIleGlnHisThAsnGlyPheValSerSerTyrIle 272
DB 484 TCTGGGTGCTCCGTC-----ACCGAGTGTATTATCTCTGCTCCATC 525

```

```

BF859163
LOCUS BF859163 839 bp mRNA linear EST 19-JAN-2001
DEFINITION 963001C10.Y3 C. reinhardtii CC-1690, Stress condition I, normalized
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF859163
VERSION BF859163.1 GI:12249254
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 839)
Grossman,A., Davies,J., Federpiel,N., Harris,E., Hauser,C.,
Lefebvre,P., McDermott,J.P., Shrago,J., Sillflow,C. and Stern,D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; Project phase 3
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source location/Qualifiers
1..839
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_1ib="C. reinhardtii CC-1690, Stress condition I,
normalized, Lambda zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
zap clones by superinfection with Exsist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) genome Research 6: 791-806."
BASE COUNT 215 a 268 c 241 g 114 t
ORIGIN
Alignment Scores:
Pred. No.: 0.326 Length: 839
Score: 124.50 Matches: 41
Percent Similarity: 32.05% Conservative: 9
Best Local Similarity: 26.28% Mismatches: 53
Query Match: 7.46% Indels: 53
Gaps: 4
US-10-018-706-2 (1-322) x BF859163 (1-839)
QY 44 GlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIle 63
DB 67 GGTGGGGCGCGCTGAACTAGTGAATCCCGGGGCTG----- 105
QY 64 ThrAspSerGlnGlyValProAsnArgTyGlnValGlnGlyAspThrValSerLys 83
DB 106 -----CAGAAATTCGGACAGAGGACCATATATGACAACTAGC----- 144
QY 84 IleAlaGlnArgTyGlyLeuAsnTyrArgGluIleGlyHisIleAsnAsnLeuAsnSer 103
DB 144 ----- 144
QY 104 SerTyrThrIleTyrThrGlyGlnTyrLeuThrLeuTyrSerGlyAspLeuLysValArg 123
DB 145 -----GGGACAGATGGGTCTTGGCTGGCAAGACCGAGCTGAAGTGGCG 189

```

Oy	124	GluatgSerIleSeSeSglYalaaenhrAlahIstnrProSeProValAlaValGln	143
Dd	190	AACAAAAAGCtAgCTCGgtTCttCCGCCctCGccCGCAcCTGTCTCCAAcCACAGTGCNA	249
Oy	144	SerSerArgProProVal-----GlnGlnHiaPro	153
Dd	250	CAGCAGACGAGCGCTCTCCGAAGAAGTCGGGCGAGCATCAAGCCAGGACGACGACT	309
Oy	154	AlaValGlnlyProThnProProValValValValylslyProThnProthrProPro	173
Dd	310	CCTGTGCCCCCGCCATCTCCACCG-----CCCAGCTTAAGCTTGCTCTCTCT	357
Oy	174	ValValGlnGlnProAlaproValAlaproProValThrGlnAlapro	189
Dd	358	GCGCAGCCCCGGCGCCGCTCGGCTCCAGAGTGTAAGAGCTATCCCG	405
RESULT 16	BH705150/c		
LOCUS	BOMKT37TF	BO_2_3_KB Brassica oleracea genomic clone BOMKT37,	DNA
DEFINITION	BOMKT37TF BO_2_3_KB Brassica oleracea genomic clone BOMKT37,		DNA
ACCESSION	BH705150		
VERSION	BH705150.1	GI:18787488	
KEYWORDS	GSS.		
SOURCE	Brassica oleracea.		
ORGANISM	Brassica oleracea. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. Town,C.D., Van Aken,S., Uteirback,T. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea Unpublished (2001) Other_GSSes: BOMKT37TR Contact: Chris Town TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdcowm@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TF Class: sheared ends.		
FEATURES	source	location/Qualifiers	
	1..285	/organism="Brassica oleracea" /strain="TO1000DH3" /db_xref="taxon:3712" /clone="BOMKT37" /clone_lib="BO_2_3_KB" /note="Vector: pHDSt; Site 1: BatXI; 2-3 kb sheared genomic DNA inserted into pHDSt using BatXI linkers"	
BASE COUNT	27 a	26 C	143 g 89 t
ORIGIN			
Alignment Scores:	Pred. No.:	0.0767	Length: 285
	Score:	124.00	Matches: 27
	Percent Similarity:	50.94%	Conservative: 0
	Best Local Similarity:	50.94%	Mismatches: 24
	Query Match:	7.43%	Indels: 2
	Dd:	17	Gaps: 1
US-10-018-706-2 (1-322) x BH705150 (1-285)			
Oy	136	ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHieProAlaVal	155
Dd	163	ACACCAAATCCACCGTTGTAAAGCCACCAACAACACCAACAACCAACTCCACAGTGGT	104
Oy	156	GlnlysProThr-----ProProValValValValyslysProThnProthrProPro	173
Dd	103	ACGCCAACCAACCGCCCGCCACCTGTATTAAAGCCACCAACAACCAACAACCAACCG	44

cy	174	ValValgIngnInPcoAlaProValAlaProProValThr	186
db	43	GTCTTAACGCCACCGACACCAACCCCTCTGTGTGAACA	5
RESULT 17			
LOCUS	A2187441/c		
DEFINITION	SP_1009_A2_D09_SP6 Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1009 Col=18 Row=G, DNA sequence.		
ACCESSION	A2187441		
VERSION	A2187441.1		
KEYWORDS	GSS.		
SOURCE	Strongylocentrotus purpuratus.		
ORGANISM	Strongylocentrotus purpuratus		
REFERENCE	Eukaryote; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.		
AUTHORS	1 (bases 1 to 735) Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Pouetka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.		
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)		
MEDLINE	20402566		
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1009 row: G column: 18 Seq primer: SP6 Class: BAC ends High quality sequence stop: 735.		
FEATURES			
Source	Location/Qualifiers 1..735 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="Plate=1009 Col=18 Row=G" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BACs; 6; BAC clones in E-Cell DH10B"		
BASE COUNT	178 a 71 c 311 g 175 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	0.299	Length:	735
Score:	124.00	Matches:	41
Percent Similarity:	45..99%	Conservative:	22
Best local Similarity:	29..93%	Mismatches:	53
Query Match:	7..43%	Indels:	21
DB:	17	Gaps:	6
US-10-018-706-2 (1-322) x A2187441 (1-735)			
cy	80	ThyValSerIysIleAlaGlnArgTyGlyLeuAsn-----TTrArgGluIleGlyHis	97
db	701	ACCAATATCTGCAGCTCACTTAAGCGTAAACCTCTCTCTCTCTCCACCTCCATACCCCA	642
cy	98	ILeAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeuThrLeuTrpSer	117
db	641	CTTAAATCACTAC--TCTTCTTACACA-----	618
cy	118	GLYAspLeuLysValAlaArgGluArgSerIleSerSerGlyValAsn-ThrAlaHisThrPr	137
db	617	---CCCTCCGTCGTGCACTTAATACACTGATCCACCACTCTCAACCCCAACACCC	561

```

QY 137 oSerProValAlaVal-----GlnSerSerArgProProValGlnGlnHisProAlaVal 155
DB 560 CCTCTCCCTCCAGCCCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 501
QY 155 GlnLysProThrProProValValValLysLysProThrProThrProProVal-- 174
DB 500 CCTCTCCCTCCAGCCCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 441
QY 175 -ValGlnGlnProAlaProValAlaProProValThrGlnLysLysProAlaThrGlyse 194
DB 440 CCTCTCCAGTTCCTCCCTCAAGATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 381
QY 194 rSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProVal 210
DB 380 CTCTC-----CTGCATTTCTCTCCACCCACTTCTCCACTGTCACTCCCTT 338

RESULT 18
LOCUS BM913728 1252 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT 6612844 NIH_MGC_98 Homo sapiens CDNA clone IMAGE:5477423
5', mRNA Sequence.
ACCESSION BM913728.1 GI:19364107
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1252)
AUTHORS NIH-MGC, http://mgi.nci.nih.gov/
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: L1CM1995 row: 9 column: 24
High quality sequence scop: 398.
Location/Qualifiers
1. 1252
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5477423"
/tissue="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10b (phage-resistant)"
/notes="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: This is a NIH_MGC library." 1 others
BASE COUNT 218 a 568 c 230 g 235 t
ORIGIN

Alignment Scores:
Pred. No.: 0.644 Length: 1252
Score: 124.00 Matches: 56
Percent Similarity: 32.77% Conservative: 22
Best Local Similarity: 23.53% Mismatches: 84
Query Match: 7.43% Gaps: 9
DB: 14

US-10-018-706-2 (1-322) x BM913728 (1-1252)
QY 21 IlshPheGlyValIleThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40

169 ATATTGGCATTTTCAGTTCTGTACGCATC-----TGCCGGTTGCAGCCACCCCTT 222
QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySer 60
DB 223 ACTTCACAGGATGC---AGCTGCTCACTTGGGGGCACTGGCTCTTAGGTTT----- 273
QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValLysGlnGlyAspThr 80
DB 274 -----AACGAT 279
QY 81 ValSerLysIleAlaGlnArgTyrGlyLeuAsnThrPargGluIleGlyHisIleAsnAsn 100
DB 280 GTCAACAGTGTAGTTTAAATGAGCCGTTAGTGTCTC-----TATTGCATATAT 330
QY 101 -LeuAsnSerSerTyrThrIleTyrThrGlyGlnThrPheThrLeuThr----- 116
DB 331 GTTAGGACATTATATGATTTTCCACGACAGTAC-----ACCATGTGGCTTAGGTAG 384
QY 117 -SerGlyAspLeu-----LysValArgGluArgSerIleSerSerGly 130
DB 385 CAGCGGGCCACTTGTATTCATCTTGACAGACTTATTGCCCCCGCTCCGCCACCC 444
QY 121 -----LysValArgGluArgSerIleSerSerGly 130
DB 445 CTCCCTCCCTCACCCGACGACGCGCCCATCTCCCTCCCTCCCTCCCTCCCTCCCTCC 504
QY 130 yValaenThrAlaHisThrProSerProValAlaValGlnSer----- 145
DB 505 CAGCGGCACACACACACACACCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 564
QY 146 -----ArgProProValGlnGlnHisProAlaValGlnLysProThrProPro 162
DB 565 CCACGCTTACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 624
QY 162 lValaValLysLysProThrProThrProThrProProValAlaGlnGlnProAlaProVal 182
DB 625 CTCTGACATCTCCGCGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 684
QY 182 aProProValThr-GlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgT 202
DB 685 CCGCGCCCTCCACCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 744
QY 202 yProVal-----GlyAlaThrAsnProVal 210
DB 745 CTCTCCCTCAGATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 794

RESULT 19
LOCUS BE240905 538 bp mRNA linear EST 16-OCT-2000
DEFINITION S650423 Sueda salsa ZAP cDNA library Sueda maritima subsp. salsa
CDNA, mRNA sequence.
ACCESSION BE240905.1 GI:9057090
VERSION BE240905
KEYWORDS
SOURCE EST.
ORGANISM Sueda maritima subsp. salsa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Chenopodiaceae; Sueda.
REFERENCE 1 (bases 1 to 538)
AUTHORS Zhang, L., Ma, C.-L., Wang, P.-P., Sun, Y.-F., Zhao, Y.-X. and Zhang, H.
TITLES Expresed sequence tags from a halophyte Sueda salsa cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Hui Zhang
Key Laboratory of Plant Stress Research
The Biology Department of Shandong Normal University
No. 88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC
Tel.: (86)531-2860864
Fax: (86)531-2866954
Email: zhangh@sdu.edu.cn.
Location/Qualifiers
1. 538
FEATURES
source

```


Oy 300 GlyAlaIalaLeuPheArgIleSerArganGlyValTYrValASPPro 317
 |||||
Db 603 TCGCACACGCCTGTTCAGTTTGAAGAAGAGCATTACCGCTTTGAGGGCGGATCC 656
 |||||

RESULT 21
LOCUS BUZ20947/c
DEFINITION BUZ20947 Y. Ogihara unpublished cDNA library, Wh Triticum aestivum
ACCSSION CDBA clone whznho9 3', mRNA sequence.
VERSION BUZ20947
KEYWORDS BUZ20947.1 GI:19959118
SOURCE EST.

ORGANISM bread wheat.
Triticum aestivum
Euryotia; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 689)
Ogihara,Y. and Murai,K.
Expressed genes in Triticum aestivum
Unpublished (2002)

JOURNAL Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel.: 81-559-81-6855
Fax1: 81-559-81-6855
Email: tshin@genie.nig.ac.jp.
location/Qualifiers
1..689

FEATURES
SOURCE

BASE COUNT	160 a	157 c	185 g	187 t
ORIGIN				
Alignment Scores:				
Pired. No.:	0.336		Length:	689
Score:	123.00		Matches:	58
Percent Similarity:	38.91%		Conservative:	27
Best Local Similarity:	26.61%		Mismatches:	86
Query Match:	7.37%		Indels:	47
DB:	13		Gaps:	10

US-10-018-706-2 (1-322) x BJ220947 (1-669)

QY	68	GLYVALPRQASNAKRGTYRGINVALVLSGNGINLYASPRTHVALSERLYSILEALGINAG	87
Db	689	GGGGTCCCAAT-----GTCAGCCCATACATACAGAAAGGCTCTCTACAGG	639
QY	88	TYRGYILEUANTPRARGINUIEGYLHISILEASNALEUANSERSEYRTHYLE	107
Db	638	ATGGGCTC-----ACCAAA	624
QY	108	TYRTHCYGLMTRPLEUTHIRLEUTYPSERGLYASPLEUYSVALARGINARGSERILE	127

[illegible]

RESULT	22
LOCUS	AU268480
DEFINITION	500 bp mRNA EST 10-MAY-2002
ACCESSION	AU268480
VERSION	AU268480
KEYWORDS	sequence.
SOURCE	AU268480.1 GI:20527278
ORGANISM	EST.
REFERENCE	Dicystostelium discoideum. Dicystostelium discoidium. Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium. 1 (bases 1 to 500) Unishihara,H., Morio,T., Saito,T., Koriiki,E., Ochiai,H., Maeda,M., Takeuchi,I., Kohara,Y. and Tanaka,Y. Population analysis of cDNAs from unicellular and multicellular stages of Dicystostelium discoidium Unpublished (2002)
TITLE	Contact: Hideko Unishihara Institute of Tsukuba Sciences University of Tsukuba 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan Tel.: 81-298-53-4664 Fax: 81-298-53-6614 Email: hideko@obi.tnkuha.ac.jp.
JOURNAL	location/Qualifiers
COMMENT	1 500
FEATURES	location= "dicystostelium discoidium"
SOURCE	

```

BASE COUNT      212 a      151 c      50 g      85 t      2 others
ORIGIN
Alignment Scores:
Pred. No.:      0.235      Length:      500
Score:          122.50     Matches:      41

```

Percent Similarity:	43.75%	Conservative:	8
Best Local Similarity:	36.61%	Mismatches:	3
Query Match:	7.34%	Indels:	3
DB:	9	Gaps:	7

US-10-018-706-2 (1-322) x AU268480 (1-500)

Qy	133	ThlAlhIghrProSerProValAlaVal	-----GlnSer	145	
Db	73	ACAACAACAACCTCCAACTCCAGCTCAAGTACCCAGCTGTACCAACAACAACATTCA	132		
Qy	146	ArgPro---	ProValGlnGlnHisProAlaValGlnIlySProThrPro---	ProValVal	163
Db	133	AAACCGACCCCACTGTGTAGACGCCCACTGTGTAGCAACCCCAACTCCAGCCCACTGCTT	192		
Qy	164	ValValIlyIyIySProThrProThrPro---	ProValIvalGlnGlnProAla-	179	
Db	193	GCAACCCCAACCCCACTCCAACTCCAACTCCAGTTACCAACCAATCCAACTGTAGAGTT	252		
Qy	180	---ProValAlaIaProProValAlaThrGluAlaIaProPhe	-----AlaThrGlySerSer	195	
Db	253	GACCACTCAAGTGGACCACTTACAGCTGCTGCTGTGTGGCCGCTGACAGGATTTGCA	312		
Qy	196	GlyValMetGlnPheArgTrpProValGlyAlaThrAsnProValValaIaArgIyPheGly	215		
Db	313	GGTATTCAA-----GGT	324		
Qy	216	ThrAlaThrValAlaGlySerThrValaThrSerAsn	227		
Db	325	ACAACCTACATCTGCAACAACAACAACAACAACAAC	360		

RESULT	23
LOCUS	AG130154
DEFINITION	Pan troglodytes DNA, clone: PTB-141N09.R., genomic survey sequence.
ACCESSION	AG130154
VERSION	AG130154.1 GI:16659319
KEYWORDS	GSS.
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1tb:PTB Chimpanzee Male

REFERENCE
AUTHORS
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of library PBs
Unpublished
2 (bases 1 to 722)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.

```

Sequencing: M13Rev
LIBRARY
  Vector      : pKS145
  R.Site 1    : SacI
  R.Site 2    : SacI
  location/Qual1
  1. . 722
FEATURES
  source

```

	clone_1b="PTB Chimpanzee Male BAC Library
BASE COUNT	166 a 397 c 89 g 65 t 5 others

Alignment Scores:	
Pred. No.:	0.339
Score:	122.50
Percent Similarity:	48.28%
Best Local Similarity:	35.63%
Query Match:	7.34%
Ds:	17
Length:	72
Matches:	31
Conservative:	11
Mismatches:	22
Indels:	4
Gaps:	23

US-10-018-706-2 (1-322) X AG130154 (1-722)

QY	125	ArgSerIleSerSerGlyValAsnThrAlaHis	-----	135
Db	138	CGACACGTGTCCAGAAAGCTCAACACACCCAC	CCCTTGGAGCTGCCAAGGCCCCACCC	197
QY	136	-----ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisPro	-----	153
Db	198	CACACCAACCCCGCCGACGACTGTGCTGGCGCCACCAACACCCCCCA	-----CCGCACCCCG	251
QY	154	-----AlaValGlnIlybProThrProProValValValLysLysPro	-----	168
Db	252	CCCGCGCACCGCAGCGGCCCCCGACCCACACCCACCCCGCA	-----CCC	296
QY	169	ThrProThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAla	-----	188
Db	297	ACCCTGACTCCAGCAGCAGCTGAGCAGCAGTATGCCCCCCCCCCCCGACACCCACGACA	-----	356
QY	189	ProPheAlaThrGlySerSer	195	
Db	357	CCCCCCCCCACCCTGCTCG	377	

LOCUS	DEFINITION	ACCESSION	VERSION
AG054509	1363 bp DNA linear	GSS 02-NOV-2001	
AG054509	Pan-troglodytes DNA, clone: PTB-040D10.F, genomic survey sequence.		
AG054509.1	GI:16591952		

REFERENCE	TITLE
AUTHORS	JOURNAL
	REFERENCE
	AUTHORS
1	
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,	
Tsokaki, Y., Watanabe, H., and Sakaki, Y.	
BAC end sequences of Library PB	
Unpublished	
2 (bases 1 to 1363)	
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,	
Tsokaki, Y., Watanabe, H., and Sakaki, Y.	

COMMENT
Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

```
FEATURES
source
location/Qualifiers
1..722
/organism="Pan troglodytes"
/db_xref="taxon:9538"
/clonetype="PTB-141N05.R"
/sex="male"
/cell_type="lymphoblast"
```



```

/clone="PTB-040D10.F"
/sex="male"
/cell_type="lymphoblast"
/clone_id="PTB Chimpanzee Male BAC Library"
BASE COUNT      266 a      413 c      442 g      219 t      24 others
ORIGIN

```

Alignment Scores:

```

Pred. No.:      0.995      Length:      1163
Score:          122.50      Matches:      69
Percent Similarity: 34.38%      Conservative: 41
Best local Similarity: 21.56%      Mismatches: 141
Query Match:      7.34%      Indels:      69
DB:              17      Gaps:      15

```

US-10-018-706-2 (1-322) x AG054509 (1-1363)

```

Qy 36 LysProThrTyraSerThrSerGlySerGlySerHisArgThrSerGlySerGly 55
Db 248 AAACCGACGAGAAAAAGAAAGAAACATTAATAAAGAGAACCGGAGGAGCGCGC 307
Qy 56 LeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValPro--AaaArgTyrgln 74
Db 308 CGC-----GGCGGCGGAGGAGGTGCGGAGAACCGCGCGCACTCGCGGCGCATCGCG 361
Qy 75 ValIleGlnIleAspThrValSerIleAlaGlnArgTyrgIleuAsnTrpArgIu 94
Db 362 GAGATGACGCGCGCGCGCGCGACCTCGCATGACTCGCGCGCAGTGTGCGCGCG 421
Qy 95 IleGlyHisIleAsnLeuAsnSerSerTyThrIleTyThrGly-----GlnTrp 112
Db 422 CGCGGCTTATATAGGATACGCGCTTCTCATCATGCTTATGCTGAGTATGCTGTG 481
Qy 113 LeuThrLeuTrpSerGlyAspLeuIysValArgGlnAspSerIleSerGly----- 130
Db 482 TCACCTCCCTTGCTCACTATTCGTCGTCACGTCGTCACCATTAACCGGACGCGC 541
Qy 131 -----ValaThrAlaIleThrProSerProValAlaGlnSerSerArg 146
Db 542 GATCAATACGTATGTGTGCTGCGGCGGTGCGGCGCGCGCGCGCGCGCGCAT 601
Qy 147 ProProValGlnGlnHisProAlaValGlnLysProThrProProValValValLys 166
Db 602 GCTTCG-----CCCGAGTTCGCGTCCGCGTTCACCGCATTCCTGTTCG 649
Qy 167 LysPro--ThrProThrProProValAlaGlnGlnProAlaProValAlaProProVal 185
Db 650 TTCCCTCATTAACCGCGCGCGGTACTATGACCATCAACCAACCAACCGCGAATC 709
Qy 186 ThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrgProValGly 205
Db 710 AATACAGCTGACGACCGCGGTGTTCGCA--CGCGGTGCGGACCTCCTGCGAC 766
Qy 206 AlaThrAsn-----ProValAlaArgArgPheGlyThrAlaThrAlaIleGlySerThr 223
Db 767 GCCTCGGTCATCAACCGCTTATGTCGCAATACGATGCGGCTTACCTCGGAGAGCGCT 826
Qy 224 ValThrSerArgnGlyMetTrpPheSerGlyArgAspGlyValPheIleAsnIleAsn 243
Db 827 -----TCCCTCTCGTTCGATCCGCGAC 847
Qy 244 AlaGlyThrValIleGlnAla-----AspHisAsnMetAspGly 256
Db 848 TGCAGTCCGTCGAGCTTCTTATGCGAGAGAGCGCGCGCGCGCGCGCGCGCG 907
Qy 257 AlaSerIleValIleGln----- 262
Db 908 GCGTCGACTGCTTGGCTGTCTTACCGCGTCAAACTCGGCGCGCATGTTTCACTGT 967
Qy 263 -----HisThrArgnGly-----PheValSerSerTyrgIle----- 272
Db 968 TGTACATCTGACGATCGACGAGCGGCGGTGCTTCGCGCATACAGATCCGACGCTAA 1027

```

```

Qy 273 -----HisIleLysAspAlaGlnValIleThrGlyAspThrValArgThrGlyGlnArg 290
Db 1028 GATCGGCGCTTACGACCGACCGACCGCGTGG-----ACAGCANAAATCACTCAGCGT 1078
Qy 291 IleAlaSerMetLysArgnGlnProSerGlyValAlaIleuPheGlnPheArgIleSerArg 310
Db 1079 TGTGAGCGCGCATATATAT--CGACGCGCGCGCGCTGTGTGTGCGGATACGCTAGA 1135

```

RESULT 25

```

BM109778      450 bp      mRNA      linear      EST 26-NOV-2001
LOCUS      EST557314 potato roots Solanum tuberosum cDNA clone cPRO5C4 5' end,
DEFINITION      mRNA sequence.
VERSION      BM109778
KEYWORDS      BM109778.1 GI:117070945
SOURCE      EST.
ORGANISM      potato.

```

```

REFERENCE      Solanum tuberosum
AUTHORS      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
      Asteridae; euasteride I; Solanales; Solanaceae; Solanum.
      1 (bases 1 to 450)
      Van der Hoeven,R., Sun,H., Karameyeva,S.A., Teai,J., Van Aken,S.,
      Utecherback,T., Chiemingo,A., Bougri,O., Buell,C.R., Romling,C.,
      Tankley,S. and Baker,B.
      Generation of ESTs from potato roots
      Unpublished (2001)
      Contact: Research Genetics, Libraries Division
      Tel: 1-800-711-6195
      Email: cdnareegen.com
      For clone info: please contact Research Genetics, Libraries
      Division tel 1-800-711-6195, email cdnareegen.com
      Seq primer: T3

```

FEATURES

```

source      1..450
      location/Qualifiers
      /organism="Solanum tuberosum"
      /cultivar="Kennebec"
      /db_xref="taxon:4113"
      /clone="cPRO5C4"
      /clone_lib="potato roots"
      /tissue_type="roots"
      /dev_stage="in vitro grown stem cuttings"
      /lab_host="SOLR"
      /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
      XhoI; supplier: Cornell University, Tankley lab;
      sequencing: The Institute for Genomic Research. Roots were
      isolated from in vitro grown stem cuttings on CM medium.
      Roots were isolated two weeks after placing the stem
      cuttings from in vitro grown plants on medium."
BASE COUNT      125 a      190 c      30 g      105 t
ORIGIN

```

Alignment Scores:

```

Pred. No.:      0.225      Length:      450
Score:          122.00      Matches:      31
Percent Similarity: 48.19%      Conservative: 9
Best local Similarity: 37.35%      Mismatches: 31
Query Match:      7.31%      Indels:      12
DB:              13      Gaps:      5

```

US-10-018-706-2 (1-322) x BM109778 (1-450)

```

Qy 136 ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal 155
Db 160 ACTTCAACCAACCACTTATGTC-----CATCAACCAATCACTCAAAACCA----- 204
Qy 156 GlnLysProThrProProValVal-----ValValLysLysPro--ThrProThr 171
Db 205 CCATCAACCAACCACTTATGTTTACCCCTATGTTTATCCACCAATCAACCAACA 264
Qy 172 ProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAla 191
Db 265 CCACATATGTGTCACCTCCATTCATTCACACCAACCACTTATGTTTCCACCTTGTTC 324

```


Qy 192 ThrGlySerSerGlyValMetGlnPheArgTyr-----ProValGlyAlaThrAsn 208
 Db 325 CCAGTCTCCGGGTGTATATACACACCGCTTACGCAAGTCTCCGGGTGTATCTCA 384
 Qy 209 ProValVal 211
 Db 385 CCCATAATT 393
 RESULT 26
 AG159588/c 1140 bp DNA linear GSS 09-JAN-2002
 LOCUS DEFINITION Pan troglodytes DNA, clone: RP43-025C14.TU, genomic survey
 sequence.
 ACCESSION AG159588
 VERSION AG159588.1 GI:16689266
 KEYWORDS GSS.
 SOURCE Pan troglodytes male lymphocytes DNA, clone: RP43-025C14.TU.
 ORGANISM Pan troglodytes
 REFERENCE Male BAC Library clone: RP43-025C14.TU.
 AUTHORS Pan troglodytes
 TITLE Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Primates; Catarrhini; Homnidae; Pan.
 JOURNAL Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Tokoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of Library RPCI-43
 Unpublished
 2 (bases 1 to 1140)
 DIRECT SUBMISSION
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimbee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
 end was generated during the R&D process and may have higher chance
 of clone tracking errors.
 COMMENT PRIMERS
 Sequencing: TU
 LIBRARY
 Vector : pBACe3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI.
 FEATURES
 location/Qualifiers
 1..1140
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="RP43-025C14.TU"
 /sex="male"
 /cell_type="lymphocytes"
 /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
 BASE COUNT 34 a 347 c 476 g 282 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.949 Length: 1140
 Score: 121.50 Matches: 47
 Percent Similarity: 41.45% Conservative: 16
 Best Local Similarity: 30.92% Mismatches: 58
 Query Match: 7.28% Indels: 31
 DB: 17 Gaps: 7
 US-10-018-706-2 (1-322) x AG159588 (1-1140)
 Qy 137 ProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
 Db 1121 CCGGGGCCACCC-----CCCGCCGGAGAACCCCGCACCGCCCGCCCGC 1080
 Qy 157 LysProThrProProValAlaValAlaValLysLysProThrProThr-----ProPro 173
 Db 1079 CGACCAAGCCCGCCACCTCCCGCTCTCCCGCCGGAACCAAGTCCCGGCGCGCCCGCC 1020

Qy 174 ValValGlnGln-ProAla---ProValAlaProProValThrGluAlaProPheAlaTh 192
 Db 1019 CAGATAGGGGCAACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 960
 Qy 192 rGlySer-----SerGlyValMet-GlnPheArgTyrProValGlyA 206
 Db 959 GGGCAAAACATCAGCCACCCACCGCGCCGACACGACCGCGCCGCAACAAAGGCG 900
 Qy 206 lathraenProValAlaArgPheGlyThrAlaThrValAlaGlySerThr---ValT 225
 Db 899 CGCGCACCCCAACCCGCAAGAACGCGACGTCACCCCGCGCGCAGACTGAGATA 840
 Qy 225 hrSerAsn-----GlyMetTrpPheSerGlyArgAspGlyAspL 238
 Db 839 CCGCAACCAACCCCGCCCGCGCGCGCACTTAAGACCGCGCGTCCGACATACAC 780
 Qy 238 euIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAla- 257
 Db 779 CGCGCAACNAACATCGCGCGCGGAGAAACAGAGCGGCAACACACCGCGCGCAT 720
 Qy 258 -----SerIleValIleGlnHis 263
 Db 719 GAAACAAACAACACAGTCCGCGCGCGCAACAT 688
 RESULT 27
 B1174098 508 bp mRNA linear EST 09-JUL-2001
 LOCUS DEFINITION OSTRF009E10_1 AD-wrmcDNA Caenorhabditis elegans cDNA similar to
 F13D12.3, mRNA sequence.
 ACCESSION B1174098
 VERSION B1174098.1 GI:14639886
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea
 ; Rhabditidae; Pelodidae; Caenorhabditis.
 1 (bases 1 to 508)
 AUTHORS Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T.,
 Jackson, C., Shin, J.T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J.,
 Lee, H., Hilti, U., Doucette-Stamm, L., Hartley, J.L., Temple, G.F.,
 Brasch, M.A., Vandenhaute, J., Lamesch, P.E., Hill, D.E. and Vidal, M.
 Open-reading-frame sequence tags (OSTs) support the existence of at
 least 17,300 genes in C. elegans
 Nat. Genet. 27 (3), 332-336 (2001)
 JOURNAL MEDLINE
 COMMENT Contact: Reboul J, Vaglio P
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 44 Binney Street, Boston, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 2425
 Email: Jerome.Reboul@dfci.harvard.edu
 Sequence tag of Gateway entry clones. The primers used were
 designed on the predicted protein encoding ORF. C. elegans ORFome
 cloning project : Contact jerome_reboul@dfci.harvard.edu or
 philippe_vaglio@dfci.harvard.edu
 POLY(A)=No.
 FEATURES
 location/Qualifiers
 1..508
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone_lib="AD-wrmcDNA"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev stage="mixed stage"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"
 BASE COUNT 141 a 148 c 104 g 115 t

ORIGIN

Alignment Scores:

Pred. No.: 0.366 Length: 508
 Score: 120.50 Matches: 54
 Percent Similarity: 34.54% Conservative: 13
 Best Local Similarity: 27.84% Mismatches: 73
 Query Match: 7.22% Indels: 55
 DB: 13 Gaps: 9

US-10-018-706-2 (1-322) x B1174098 (1-508)

QY 28 CyeleleuallaglyCyala-----SertlypProthryTyranser 41
 Db 3 TGCCTCATTTTCATTCACACGCGATTTCGTAAGAACTCGGAGAGACCTACATGTC 62
 QY 42 ThsSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaileGlySerGln 61
 Db 63 AGCAGT-----TCATATCTAGGCCCAAGTTACGAGGA----- 95
 QY 62 ValIleThrAspSerGlnGlyValProAspAlaGlyGlnValIleGlnGlyAspThrVal 81
 Db 96 -----GACCAATGGGCTCAGAAATTAATTAATTAATCAATATCAAAACAATTGG-- 143
 QY 82 SerIleAlaIleAlaGlyGlyGlyLeuAspThrArgGluileGlyHisIleAspAsnLeu 101
 Db 144 -----AATCAACAAACCAAGAAATTCGATCAATATGGA-----AACAGA 182
 QY 102 AsnSerSerGlyThrIleGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 121
 Db 183 GGTAGTCATATGATGATCTCATACGCGCGGCC--TTCAGATTATGGCCACACAGCCAG 241
 QY 122 ValArgGlnArgSerIle-SerSerGlyValAspThrAlaIleThrProSerProValAl 141
 Db 242 CTACTCCCAACCAAAATCTCATATGGA-----CGCCCAAGAACCCACCTA 286
 QY 141 aValGlnSerSerAspPro-----ProValG 150
 Db 287 CTCGCCATATTCGGGCTTACCTTCATATGGAACACCACTCACTTATCCAGCTCA 346
 QY 150 ngInIleProAlaValAlaGlnIleProThrProProValAlaVal----- 164
 Db 347 GCCAGGTCCAGGAGGTCAGCTGATATCTCCAGTGTGACCTCAGCCCAATTGGCCGCC 406
 QY 165 -----ValIleGlyProThrProThrProProValAlaGlnIleProAlaPr 180
 Db 407 TCCAGTTCCAGAAAGTTTCTAGCCAGTTCAGAGCCACCA--ATCCAGCTCCAGGCC 463
 QY 180 oValAlaProProValThrGlnAlaIleProPheAlaThrGly 193
 Db 464 TGTTCCTCCACCAACCATCTAGTCTCATATTCGAGGA 503

RESULT 28

CNS03NOR/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Human genome number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 0.926 Length: 969
 Score: 120.50 Matches: 30
 Percent Similarity: 39.29% Conservative: 3
 Best Local Similarity: 35.71% Mismatches: 30
 Query Match: 7.22% Indels: 21
 DB: 17 Gaps: 2

US-10-018-706-2 (1-322) x CNS03NOR (1-969)

QY 135 HisThrProSerProValAlaValGlnSerSerArgProProProValAlaGlnHisProAla 154
 Db 606 CACCAAGCTCTCTTTCATATGACATCAAGCCCGCCCTCTCGGAACCAACACACACTGCA 547
 QY 155 -----ValGlnIleProThrProProProValAlaValAlaValAla 166
 Db 546 CACGTGCTGCGCTGTTTGCACCAAGGATGACCTGCACCCCGCTGCACCTTACT 487
 QY 167 LysProThrProThrProProProValAlaGlnIleProAlaProProProValThr 186
 Db 486 CCACCTGCACCTTACTGCTCACCCTGATGCTCACCCTGCACCTGATGCTCACCCTGACT 427
 QY 187 GluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAla 206
 Db 426 GTAGCTTCA-----CTGTAGCTTCA 406
 QY 207 ThrAsnProVal 210
 Db 405 CTGCACTGCTGTA 394

RESULT 29

AL377972

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Human genome number estimate provided by genome wide analysis using

Unpublished 2 (bases 1 to 969)

Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,

Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and

Weissenbach, J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Unpublished 3 (bases 1 to 969)

Genoscope, Direct Submission

Submitted (12-APR-2000)

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis

genome. For more information, please take a look at

http://www.genoscope.cns.fr/tetraodon.

Location/Qualifiers

1. 969

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="040D21"

/note="Genoscope sequence ID : COBG040CB11UP1-end : T7"

BASE COUNT 215 a 198 c 398 g 151 t 7 others

ORIGIN

Alignment Scores:

Pred. No.: 0.926 Length: 969
 Score: 120.50 Matches: 30
 Percent Similarity: 39.29% Conservative: 3
 Best Local Similarity: 35.71% Mismatches: 30
 Query Match: 7.22% Indels: 21
 DB: 17 Gaps: 2

US-10-018-706-2 (1-322) x CNS03NOR (1-969)

QY 135 HisThrProSerProValAlaValGlnSerSerArgProProProValAlaGlnHisProAla 154
 Db 606 CACCAAGCTCTTTCATATGACATCAAGCCCGCCCTCTCGGAACCAACACACACTGCA 547
 QY 155 -----ValGlnIleProThrProProProValAlaValAlaValAla 166
 Db 546 CACGTGCTGCGCTGTTTGCACCAAGGATGACCTGCACCCCGCTGCACCTTACT 487
 QY 167 LysProThrProThrProProProValAlaGlnIleProAlaProProProValThr 186
 Db 486 CCACCTGCACCTTACTGCTCACCCTGATGCTCACCCTGCACCTGATGCTCACCCTGACT 427
 QY 187 GluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAla 206
 Db 426 GTAGCTTCA-----CTGTAGCTTCA 406
 QY 207 ThrAsnProVal 210
 Db 405 CTGCACTGCTGTA 394

RESULT 29

AL377972

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Human genome number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Tetraodon nigroviridis DNA sequence

TITLE
JOURNAL
COMMENT

'V. and Gamas, P.
Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
Unpublished (2000)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqef@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, CNRS-IRRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mttruncatula.html).
Location/Qualifiers

FEATURES

source

1. 365
/organism="Medicago truncatula"
/cultivar="Uemalong"
/db_xref="taxon:3880"
/clone="M1B35B07"
/clone_lib="M1B3"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
/note="Vector: pBluescript PSK, Site 1: EcoRI; Site 2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-Zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exasit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

BASE COUNT

111 a 129 c 36 g 89 t

ORIGIN

Alignment Scores:
Pred. No.: 0.253 Length: 365
Score: 120.00 Matches: 26
Percent Similarity: 63.79% Conservative: 11
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 7.19% Indels: 8
DB: 9 Gaps: 3

US-10-018-706-2 (1-322) x AL377972 (1-365)

QY 137 ProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
DB 83 CTTTCCACCATGTAATAATCA-----CCACTTATCAATCAACCACTATGTTAG 136
QY 157 LysProThrProProValAla-----ValAlaValLysProThrProThr 171
DB 137 GCACCTTCTCCCTCTAGTGAACCACTCACTATGTAATCACT---CTTCT 193
QY 172 ProProValValGlnGlnProAlaProValAlaProProValAlaPro 189
DB 194 CCTCTTGTGTAAGAAACACCACTTATCAATCAACCACTATGTAAGGACCA 247

RESULT 30

AMS86923

LOCUS AMS86923 511 bp mRNA linear EST 07-SEP-2000
DEFINITION ESTJ18546 MHAM Medicago truncatula/Glommus versiforme mixed EST
ACCESSION AMS86923
VERSION AMS86923.1 GI:7266437
KEYWORDS
SOURCE Medicago truncatula/Glommus versiforme mixed EST library.
ORGANISM Medicago truncatula/Glommus versiforme mixed EST library
Eukaryota; mixed EST libraries.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 511)
Harrison, M.J., Liu, J., Peng, H., Gonzales, M., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.
ESTs from roots of Medicago truncatula after colonization with Glommus versiforme
Unpublished (2000)
Contact: Maria J. Harrison
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401, USA
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N253964e
TIGR sequence name: MTDAV15TK
More information is available at:
http://chrystle.tamu.edu/medicago/
Seq primer: SKmod (CTA gaa cta gtc gat cc).
Location/Qualifiers

FEATURES

source

1. 511
/organism="Medicago truncatula/Glommus versiforme mixed EST library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="PMHAM-56C5"
/clone_lib="MHAM"
/tissue_type="roots colonized with Glommus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glommus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glommus versiforme. The cDNA was directionally ligated into the UniZap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-asit helper phage and propagated in XL0LR cells."

BASE COUNT

158 a 169 c 59 g 125 t

ORIGIN

Alignment Scores:
Pred. No.: 0.41 Length: 511
Score: 120.00 Matches: 26
Percent Similarity: 63.79% Conservative: 11
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 7.19% Indels: 8
DB: 10 Gaps: 3

US-10-018-706-2 (1-322) x AMS86923 (1-511)

QY 137 ProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
DB 208 CTTTCCACCATGTAATAATCA-----CCACTTATCAATCAACCACTATGTTAG 261
QY 157 LysProThrProProValAla-----ValAlaValLysProThrProThr 171
DB 262 GCACCTTCTCCCTCTAGTGAACCACTCACTATGTAATCACT---CTTCT 318
QY 172 ProProValValGlnGlnProAlaProValAlaProProValAlaPro 189
DB 319 CCTCTTGTGTAAGAAACACCACTTATCAATCAACCACTATGTAAGGACCA 372

RESULT 31

B0147993

LOCUS B0147993 650 bp mRNA linear EST 24-APR-2002
DEFINITION NR049809FL1070 Developing flower Medicago truncatula cDNA clone
ACCESSION B0147993
VERSION B0147993.1 GI:20285052
KEYWORDS
SOURCE Medicago truncatula cDNA clone
EST.

SOURCE
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 650)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzalez, R.A., Bell, C.J.,
Flores, H.R., Iman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula flower library
Unpublished (2001)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 650 Std Error: 0.00
Plate: 049 Row: 2 Column: 08
Seq primer: TCACACGGAATACCTATGAC.
Location/Qualifiers
1..650
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_1fb="NF049E09FL"
/clone_2fb="NF049E09FL"
/issue_type="Developing flowers"
/dev_stage="Developmentally pooled. Contains a mixture of
very young, developing, fully-opened flowers and flowers
in early transition into pods."
/note="Vector: Lambda Zap; cDNA was prepared from polyA+
RNA from very young, developing, fully-opened flowers and
flowers transitioning into pods. The cDNA was
directionally ligated into the Uni-Zap XR vector
(Stratagene) and packaged using the Gigapack III Gold
packaging extracts. Phagemids containing cDNA inserts were
in vivo excised from the recombinant Uni-Zap XR vector
using Exassist helper phage and the E. coli strain
XL1-Blue MRF' (Stratagene). Excised plasmids were plated
using SOR cells."

BASE COUNT 197 a 198 c 77 g 175 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 0.579 Length: 650
Score: 120.00 Matches: 26
Percent Similarity: 63.79% Conservative: 11
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 7.19% Indels: 8
Gaps: 3

US-10-018-706-2 (1-322) x BQ147993 (1-650)

QY 137 ProSeProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
Db 311 CCTTCTCCACCATGTTGTAATCA-----CCACCTTATCAATCAACCACTTATGTTAG 364
QY 157 LysProThrProProValVal-----ValValLysLysProThrProThr 171
Db 365 GCACCTTCTCCCTCTAGTGAACAACCACTCCACCTATGTAATCACTCT--CCTTCT 421
QY 172 ProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaPro 189
Db 422 CCTCTTGTGTGTAACAACCACTTATCAATCAACCACTTATGTAAGGACCA 475

RESULT 32
BI271940 710 bp mRNA linear EST 18-JUL-2001
LOCUS NF016F08FL1074 Developing flower Medicago truncatula cDNA clone
DEFINITION NF016F08FL 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 710)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzalez, R.A., Bell, C.J.,
Flores, H.R., Iman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula flower library
Unpublished (2001)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 710 Std Error: 0.00
Plate: 016 Row: 2 Column: 08
Seq primer: TCACACGGAATACCTATGAC.
Location/Qualifiers
1..710
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_1fb="NF016F08FL"
/clone_2fb="NF016F08FL"
/issue_type="Developing flowers"
/dev_stage="Developmentally pooled. Contains a mixture of
very young, developing, fully-opened flowers and flowers
in early transition into pods."
/note="Vector: Lambda Zap; cDNA was prepared from polyA+
RNA from very young, developing, fully-opened flowers and
flowers transitioning into pods. The cDNA was
directionally ligated into the Uni-Zap XR vector
(Stratagene) and packaged using the Gigapack III Gold
packaging extracts. Phagemids containing cDNA inserts were
in vivo excised from the recombinant Uni-Zap XR vector
using Exassist helper phage and the E. coli strain
XL1-Blue MRF' (Stratagene). Excised plasmids were plated
using SOR cells."

BASE COUNT 219 a 205 c 94 g 187 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 0.657 Length: 710
Score: 120.00 Matches: 26
Percent Similarity: 63.79% Conservative: 11
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 7.19% Indels: 8
Gaps: 3

US-10-018-706-2 (1-322) x BI271940 (1-710)

QY 137 ProSeProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
Db 293 CCTTCTCCACCATGTTGTAATCA-----CCACCTTATCAATCAACCACTTATGTTAG 346
QY 157 LysProThrProProValVal-----ValValLysLysProThrProThr 171
Db 347 GCACCTTCTCCCTCTAGTGAACAACCACTCCACCTATGTAATCACTCT--CCTTCT 403
QY 172 ProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaPro 189
Db 404 CCTCTTGTGTGTAACAACCACTTATCAATCAACCACTTATGTAAGGACCA 457

RESULT 33
BM457264

LOCUS	BM457264	1460 bp	mRNA	linear	EST 05-FEB-2002
DEFINITION	AGENCOURT_6411711 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583548				
ACCESSION	5', mRNA sequence.				
VERSION	BM457264				
KEYWORDS	BM457264.1 GI:18506304				
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				
TITLE	Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.				
JOURNAL	1 (bases 1 to 1460)				
COMMENT	NIH-MGC http://mgs.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cga@bbs-rt@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at: http://image.llnl.gov Plate: LILAM12346 row: m column: 21 High quality sequence start: 131 High quality sequence stop: 344. Location/Qualifiers 1..1460 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5583548" /clone_lib="NIH_MGC_92" /tissue_type="embryonal carcinoma, cell line" /host="DH10B (phage-resistant)" /note="Organ: Testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."				
BASE COUNT	233 a 669 c 208 g 350 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	1.85	Length:	1460		
Score:	120.00	Matches:	36		
Percent Similarity:	45.38%	Conservative:	18		
Best Local Similarity:	30.25%	Mismatches:	33		
Query Match:	7.19%	Indels:	32		
DB:	13	Gaps:	7		
US-10-018-706-2 (1-322) x BM457264 (1-1460)					
QY	112 TrypteuthirleutpserglYAspleuysValargGInargserIleserSerclYVal	131			
Db	432 TGGCTGCGGTGCACCGAGGAACTGTCCAAAGCCGAAAAACCA-----	476			
QY	132 AenThralaIiarnrProserProValaA-----ValGInserSerArg-----	146			
Db	477 -----CATACCCCTGCGCCCTCTCCCTCCGCGCCGTCCTCTCCACACCCCTTC	527			
QY	147 ProProValGInGInHisProValaValGInLysProThrProProValaValaValys	166			
Db	528 CCCCTCCGACACAGTTCACCGCGCCCATGCGCTCCGCCCTCCATCA-----	575			
QY	167 LysProThrProThrProProValaValGInGInProAlaProValaAlaProProValThr	186			
Db	576 -----ACACCGCGCGCCCAATATTCACACCCCGCCCCCTT---CCCCCGTTTCT	626			
QY	187 GluaIaProPhaIaThrGlySerSerSerGlyValMet-----	198			
Db	627 AATCCCGCATTC-----TCCCGCAGACTTCACACCTTCCCGCCCTTCCCC	674			
QY	199 GlnPhaIaGlyrrProValaGlyAlaThrAsnProValaValaGtaGpPhneGlyThrAla	217			

```

DB          675 CATTAACCGCTTACCCCGATACGCGCCCTCCACTCCATGCGGATTCACACATCC 731
RESULT 34
Bg607114
LOCUS
DEFINITION
WHE2489_E06_J11S Triticum monococcum early reproductive apex cDNA
library Triticum monococcum cDNA WHE2489_E06_J11, mRNA
sequence.
ACCESSION
Bg607114
VERSION
Bg607114
KEYWORDS
SOURCE
ORGANISM
Triticum monococcum.
Triticum monococcum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticeae; Triticum.
1 (bases 1 to 567)
Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia
,C.C., Kang,Y., Lazo,G.R., Miller,R., Rauech,C.J., Seaton,C.L.,
Stamova,B. and Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes - Early reproductive apex cDNA library from Triticum
monococcum
JOURNAL
Unpublished (2001)
COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacifica
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stragene SX primer.
FEATURES
Source
1..567
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/cclone="WHE2489_E06_J11"
/cclone_lib="Triticum monococcum early reproductive apex
cDNA library"
/tissue_type="Early reproductive apex"
/dev_stage="seven week-old plants"
/lab_host="E. coli XL0LR"
/notes="Vector: lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; The tissue, total RNA, and
poly(A) RNA were prepared from apex at double-ridge stage
to terminal-spikelet stage during transition from
vegetative state to flower state, a cDNA library was made,
and the cDNA clones were in vivo excised at the
University of California, Davis (V. Echenique, B. Stamova
, J. Dubcovsky). Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
```

QY 54 YGlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrG1 74
 DB 96 AGGTCTCAACCTCCGCGTACAAAGCATC-----CCCCAGAGTCTCCCGCGCGGTAC 149
 QY 74 nValys---GlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrpAr 93
 DB 150 ACCCGCCCGCCAGGTCTCAACCGCGGTACAAAGCCCAAGGTCTC-----201
 QY 93 gGluIleGlyIleAlaAsnLeuAsnSerTyrThrIleTyrThrGlyGlnTrpLe 113
 DB 201 -----201
 QY 113 uThrLeuTrpSerGlyAspLeuLysValArgGlnArgSerIleSerSerGlyValAsnTh 133
 DB 202 -----ACCGCGCGGTACAAAGCATC-----CCCCAGAGTCTCCCGCGCGGTAC 251
 QY 133 r-AlaHisThrProSerProValAla-----ValGlnSerSerArgProProValG 150
 DB 252 CTGTCCCGAAGCCCTCACCGCGGTACAAAGCCCTCACCGCGCGGTACCGG 311
 QY 150 InGlnHisProAlaValGlnLysProThrProProValAlaValIleValLysProThrP 170
 DB 312 CCGCAACTCCGCGCGGTACAAAGCCCTCACCGCGGTACCGG 365
 QY 170 rGThProProValAlaGlnGlnProAlaProValAlaProPro 184
 DB 366 CTACTCTCTCAACCTACAAAGCCCTCACCGCGCGGTACCGG 409

RESULT 35
 B0646351 909 bp mRNA linear EST 15-JUN-2002
 LOCUS AGENCOURT_8286354 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6299462
 DEFINITION 5' mRNA sequence.
 ACCESSION B0646351
 VERSION B0646351.1 GI:21770523
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 909)
 NIH_MGC http://mgi.mc.ncl.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabbs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
 http://image.llnl.gov
 Plate: LRCM2512 row: c column: 15
 High quality sequence strip: 589.

FEATURES

source

Location/Qualifiers

1..909

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_11b="NIH_MGC_100"

/tissue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="CDNA made by oligo-dt priming. Directionally cloned
into EcorI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

BASE COUNT

179 a 332 c 235 g 163 t

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	119.00	62	26	909	26	8
Percent Similarity:	37.77%	26	61%	92	53	
Best Local Similarity:	7.13%					
Query Match:	14					
DB:	14					

US-10-018-706-2 (1-322) x B0646351 (1-909)

QY 26 ThrThrcysIleLeuAlaGlyValAspSerGlnGlyValProAsnArgTyrG1 45
 DB 189 ACTTCTCAGAGACTTGAAGTCTGACAGAGGTGTATCTTACCGCTCAACCTTCTC 248
 QY 46 -----GlySerHisArgThrSerGlySer--GlyGlyLeuAlaIleGly 59
 DB 249 TCGCGCCATCAACCGCTCTATGCACTACAGAGGTCTGAGCGCTGTGCAACAC 308
 QY 60 SerGlnValIleThrAspSer-GlnGlyValProAsnArgTyrGlnValLysGlnGly 79
 DB 309 ACCCGCCAGACCAAGCCGCTTACAGAGGTCTGAGCGCGGTACAGAGTACAGGAGA 368
 QY 79 pThrValSerIleLeuAlaGlnArgTyrGly--LeuAsnTrpArgGluIleGlyHisIle 98
 DB 369 T-----GTTGGACAGCTCTCAAGGTACAGATGATCAAGATGAGATTTCCAGAGTTCG 422
 QY 99 AsnAsnLeuAsnSerSerTyrThrIleTyrThrGlnTrpLeu-----113
 DB 423 ACGAATCTCAAGAAAGATTGATGCGATGATGATGATGATGATGATGATGATGATGATGAT 482
 QY 114 ThrLeuTrpSerGlyAspLeuLysValArgGlnArgSerIleSerSerGlyValAsnTrp 133
 DB 483 TCGCTCTGGAGAGCTTCAAGAGCTCTGAGAGAGGAGCTTCAAGAGAGCTTCTTC 542
 QY 134 AlaHisThrProSer-----ProValAlaValGln 143
 DB 543 TGTTCACAGAGCTCTGCTATACAGAGCGCGGCGCTGACGCGCTTCAATCATGTTAAG 602
 QY 144 SerSerArgProProValGlnGln-His-----152
 DB 603 CCGCGGACACTCTCCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 662
 QY 153 -----ProValAlaValGlnLysProThr-----ProProVal 162
 DB 663 CGCGCCCTGCTGCGCTCTCCCGCTCCGCGCGCGCTTATGATGATGATGATGATGATGATGAT 722
 QY 162 lValValValLysLysProThrProThrProProValAlaGlnGlnProAlaProValAl 182
 DB 723 CCGGATCTCAACAGCGCGCACCGCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCT 782
 QY 182 aProProValThrGlnAlaProPheAla--ThrGlySerSerGlyValMetGlnPheArgT 202
 DB 783 CCGACTTCAACCAAGCG 211
 QY 202 yrProValGlyAlaThrAsnProValAl 211
 DB 831 TCCCCCCCCCAGGACCAACCCGTTCTC 859

RESULT 36

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

A1484465 443 bp mRNA linear EST 18-MAY-2001
 EST242475 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
 CLRD116 similar to proline rich protein, mRNA sequence.
 A1484465
 A1484465.1 GI:4379836
 EST.
 SOURCE Tomato.
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasteride I; Solanales; Solanaceae; Solanum;

REFERENCE	Lycopersicon.
AUTHORS	1 (bases 1 to 443)
TITLE	Alcala,J., Vredalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Konning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nieman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
JOURNAL	Generation of ESTs from tomato carpel tissue unpublished (1999)
COMMENT	Other_ESTs: EST242465 Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html . Location/Qualifiers
FEATURES	source 1..443 /organism="Lycopersicon esculentum" /cultivar="TA496" /db_xref="taxon:4081" /clone="cLEDH16" /clone_lib="tomato ovary, TAMU" /tissue_type="carpel" /dev_stage="5 days pre-anthesis to 5 days post-anthesis" /lab_host="XLA-Blue MRP." /note="vector: p Bluescript SK(-); Site 1: EcoRI; Site 2: XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
BASE COUNT	119 a 187 c 31 g 106 t
ORIGIN	
Alignment Scores:	
Pred. NO.:	0.507 Length: 443
Score:	118.00 Matches: 33
Percent Similarity:	48.19% Conservative: 7
Best local Similarity:	39.76% Mismatches: 31
Query Match:	7.07% Indels: 12
DB:	9 Gaps: 5
US-10-018-706-2 (1-322) x AI484465 (1-443)	
OY	136 ThrProSerProValAlAlValGlnSerSerArgProProValGlnGlnHisProAlVal 155 :::
Db	175 ACTCCAAACACCACCTATTGTC-----CATCCACAGTCATCCAAAACA----- 219 :::
OY	156 GlnlysProThrProProValAl-----ValAllyllysPro---ThProthr 171 :::
Db	220 CCATACCAACACCTCTATTGTTCACCCCCCATTTGTTATCCACCAATCACTCCAACA 279 :::
OY	172 ProProValAlGlnGlnProAlaProValAlAProProValThrGlualaProPhlea 191 :::
Db	280 CCACCTGTTGGTGACTCCATCATATTCACACACCACTATTGTCTCCACCTTTTCTC 339 :::
OY	192 ThrglySerSerGlyValMetGlnPheargtyr-----ProValGlyAlaThrAsn 208 ::: :::
Db	340 CCCATCTCGCCGCTGTATAACCAACCACTACGTGCCAAGCTCTCGATTGTACTCCA 399 ::: :::
OY	209 Provalval 211 :::
Db	400 CCCATAGTT 408 :::
RESULT 37	
LOCUS	BA460175 443 bp mRNA linear EST 18-MAY-2001
DEFINITION	ES415467 tomato developing/immature green fruit Lycopersicon
ACCESSION	BA460175
VERSION	BA460175.1 GI:9504477
KEYWORDS	EST.
SOURCE	tomato.
ORGANISM	Lycopersicon esculentum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

[illegible]

H76194
LOCUS
DEFINITION
17899 Lambda-PRU2 Arabidopsis thaliana cDNA clone 196014T7, mRNA
sequence.
ACCESSION
H76194
VERSION
H76194.1
KEYWORDS
GI:1053445
SOURCE
EST.
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosidae; eudicots; Magnoliophyta; eudicotyledons; core eudicots;
Spermatophyta; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 615)
Newman,T., deBruijn,F.J., Green,P., Keesstra,K., Kende,H., McIntosh
,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel
,E. and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRU, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 223jlc@lpm.cl.msu.edu
Seq primer: T7 dye primer.
Location/Qualifiers
1..615
/organism="Arabidopsis thaliana"
/strain="var columbiana"
/db_xref="taxon:3702"
/clone="196014T7"
/clone_1fb="lambda-PRU2"
/note="Vector: lambda Zip-lox; Site 1: Sal; Site 2: Not;
lambda PRU2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light/8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques). The vector is BRU's lambda Zip-lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dT primed cDNA. 182 t 24 others

BASE COUNT 132 a 154 c 123 g 182 t 24 others
ORIGIN

Alignment Scores:
Pred. No.: 0.902 Length: 615
Score: 117.50 Matches: 36
Percent Similarity: 42.37% Conservative: 14
Best Local Similarity: 30.51% Mismatches: 50
Query Match: 7.04% Indels: 18
DB: 14 Gaps: 4

US-10-018-706-2 (1-322) x H76194 (1-615)

QY 134 AAlaHsThProSePrProValAlaValGInSeSerArGPrProProValGInGInHsPro 153
DB 93 GCTCAAGCTCTGCAACCAACCAACCGTACTCTCTCAACCAACGAGCGCTTCTCA 152
QY 154 ----AlaValGInLysProThPro-----ProValValValValLysLysProTh 169
DB 153 GTCACGACGAGAAACCTTCAACCAATTGCTCAACGCGCTGTTCCAGTTAAACAGCAACT 212
QY 170 ProThProProValValGInGInProAlaProValAla---ProProValThGInAla 188
DB 213 CGAGCTCTACCACTTCTCAACCAACATCCCGGTGCTTCTCTCTCAAGCCGATGCT 272
QY 189 ProPheAlaThrGlySerGlyVal-MetGInPheArgTyrProValGlyAlaThrAs 208

DB 273 CTTGCTCCCGGTCCTTCCGCTGGATTAACTCAACTTCTCTCAAGCTTCCAGGCGCCAG 332
QY 208 nProValValArgPheGlyThrAla----- 217
DB 333 AACGCTCTCTCAATCAACAGTCCGCAATGGCTAACAAAGTTTTCTTGTGGG 392
QY 218 ----ThyValAlaGlySerThrValThrSerAsnGlyMetTyrPheSerGly 233
DB 393 AACAGCTTCTCCGAGGTTTATACGGGATGCTTTTGGCTTAGAGG 444

RESULT 39
BH244723/C
LOCUS
DEFINITION
ATZEC82TF ATZE Arabidopsis thaliana genomic clone ATZEC82, DNA
sequence.
ACCESSION
BH244723
VERSION
BH244723.1
KEYWORDS
GI:16921883
SOURCE
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosidae; eudicots; Magnoliophyta; eudicotyledons; core eudicots;
Spermatophyta; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 754)
Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Uteback,T.V.,
Feldblum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC T6316
Unpublished (2001)
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1144. Caution: the DNA in this BAC may be from
some non-Arabidopsis source
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..754
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="ATZEC82"
/clone_1fb="ATZE"
/note="Vector: pHO82; Site 1: BstXI; 2-3 kb sheared BAC
DNA inserted into pHO82 using BstXI linkers"

BASE COUNT 124 a 199 c 355 g 76 t
ORIGIN

Alignment Scores:
Pred. No.: 1.21 Length: 754
Score: 117.50 Matches: 32
Percent Similarity: 48.28% Conservative: 10
Best Local Similarity: 36.78% Mismatches: 40
Query Match: 7.04% Indels: 5
DB: 17 Gaps: 3

US-10-018-706-2 (1-322) x BH244723 (1-754)

QY 134 AAlaHsThProSePrProValAlaValGInSeSerArGPrProProValGInGInHsPro 153
DB 402 GCTGCGCCGCGCGCGCGCTGCTCACTTCACTTCCCGCATATATGTCGCGGCC 343
QY 154 AAlaValGInLysProThPro-----ValValValValValLysLysProTh 171
DB 342 GCATACACGCTCTGCG 283
QY 172 ProProValValGInGInProAlaProValAlaProProValThGInAlaProPheAla 191
DB 282 GCCCATATGATGCG 229
QY 192 ThnGlySerSerGlyValMetGInPheArgTyrProValGlyAlaThrAsnProValVal 211

Db 228 TCCTGCTCTCTCCGCGCCATCCACATCCCGTT---GCCCGAGCCCGCATGATG 172
 Qy 212 ArgArgPheGlyThrAlaThr 218
 Db 171 CCCCGTGCACCCGCCACC 151

RESULT 40

U74096

LOCUS U74096 329 bp mRNA linear EST 14-APR-1997

DEFINITION ATU74096 NaCl-treated Arabidopsis subtraction library Arabidopsis thaliana cDNA clone OS001, mRNA sequence.

ACCESSION

U74096

VERSION U74096.1 GI:1932828

KEYWORDS

EST.

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 329)

REFERENCE

Pih,K.T., Park,J.M., Jang,H.U., Kang,S.G., Piao,H.L. and Hwang,I.

TITLE

EST of salt inducible mRNA in Arabidopsis thaliana

Unpublished (1997)

COMMENT

Contact: Kyeong Tae Pih

Department of Plant Molecular Biology
 Gyeongsang National Univ., Plant Molecular Biology and

Biotechnology Research Center
 Jinju, Kyounghnam 660-701, Korea

Tel: 82-591-751-5193
 Fax: 82-591-759-9363

Email: inhwang@nongae.gsnu.ac.kr.
 Location/Qualifiers

FEATURES

source

1..329
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="OS001"
 /clone_lib="NaCl-treated Arabidopsis subtraction library"
 /dev_stage="7 day-old"
 /note="Subtracted cDNA library from salt(NaCl)-treated whole plants"

BASE COUNT 91 a 130 c 46 g 53 t 9 others

ORIGIN

Alignment Scores:

Pred. No.: 0.407 Length: 329
 Score: 117.00 Matches: 26
 Percent Similarity: 51.72% Conservative: 4
 Best Local Similarity: 44.83% Mismatches: 24
 Query Match: 7.01% Indels: 4
 DB: 14 Gaps: 2

US-10-018-706-2 (1-322) x U74096 (1-329)

Qy 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
 Db 53 GTGCTAACACGCGCAACGCACTCCACCGTCATT-----ACACCAACCAACACNA 103
 Qy 151 GlnHisProAlaValGlnLysProThrProProValValValLysLysProThrPro 170
 Db 104 ACTGCACCCGTCATTACACCAACCAACCAACCAACCCGTCGTGACACCTCCCAACACCA 163
 Qy 171 ThrProProValValGlnGlnProAlaProValAlaPro---ProValThnGln 187
 Db 164 ACACCAACCCGTAGTACACCAACCAACCAACCAACCTCTACGCTATTCCAGAG 217

Search completed: July 3, 2003, 21:28:26
 Job time : 1155 secs

THIS PAGE BLANK (USPTO)